

An innovative approach to inventory and monitoring of natural resources in the Mexican State of Jalisco

Robin M. Reich · Celedonio Aguirre-Bravo ·
Martin A. Mendoza Briseño

Received: 31 May 2007 / Accepted: 12 November 2007 / Published online: 18 December 2007
© Springer Science + Business Media B.V. 2007

Abstract This paper discusses the statistical design and estimation processes developed for the assessment of key land resources relevant to questions of their condition and change in the State of Jalisco, Mexico. Some initial results of the first phase of Jalisco's Natural Resource Inventory and Monitoring Program (IMRENAT) conducted in 2006 are presented and discussed. Since this is a relatively new approach for an inventory and monitoring program of this magnitude, designed from the beginning specifically for integration with orbital satellite data, it is anticipated that changes will have to be made over time to improve conceptual and operational design aspects. Alternative remote sensing capabilities to further improve local information, more sophisticated analysis to satisfy the requirements of officials as they

become more aware of the capabilities of the system, possible improvements in change assessments over time, and opportunities to focus on specific interesting changes will be assessed. As this study indicates, current and future information technology advancements provide a solid prospect for the development and application of integrative sampling strategies to support sustainability science and management processes.

Keywords Bioclimatic zones · Domain sampling · Tree species diversity · Spectral and spatial variability

Introduction

A sound statistical strategy is fundamental to the design of scientifically credible inventory and monitoring programs. What to measure, where and when, are issues that are directly related to the goals and objectives of what is expected to be accomplished and reported by any ecosystem resource inventory or monitoring assessment program. Each inventory or monitoring program has its own particular constraints that define its conceptual and operational framework and expected outputs. Inventorying and monitoring are two distinctive interconnected processes whose statistical designs may not be compatible. Making a clear distinction if it is an inventory or a monitoring process is central to the decision of what statistical strategy is more appropriate and why. Technical issues of geographical scale, spatial and temporal

R. M. Reich (✉)
Department of Forest, Rangeland and Watershed,
Colorado State University,
Fort Collins, CO 80523, USA
e-mail: robin@warnercnr.colostate.edu

C. Aguirre-Bravo
Rocky Mountain Research Station, USDA Forest Service,
Fort Collins, CO 80525, USA

M. A. Mendoza Briseño
Colegio de Postgraduados, Campus Montecillos,
Texcoco, Mexico

M. A. Mendoza Briseño
IMRENAT, SEDER-FIPRODEFO,
Jalisco, Guadalajara, Mexico

resolutions, sampling frames, sample size and distribution, plot configuration (i.e., size and shape), accuracy, precision, and estimation strategies are key statistical aspects that must be clearly specified when designing inventory and monitoring programs. For example, a different statistical strategy might be recommended if the program objectives are strategic, tactical, or operational for local applications. These and other considerations suggest that designing either an inventory or a monitoring program is not a straight forward compatible process. Clearly, a variety of questions must be sorted out in order to recommend the most appropriate statistical strategy to meet the goals and objectives of the desired inventory or monitoring program.

How to account for the above considerations when designing a statistical strategy is a challenging process of multi-relational dimension. The Natural Resources Inventory and Monitoring Program (IMRENAT) of the Mexican State of Jalisco provides a wealth of tangible examples of the complexities involved in the process of designing a defensible statistical strategy for monitoring and assessment of terrestrial ecosystem resources at multiple scales and organizational levels. Initially, IMRENAT was conceived as a learning process for developing an *integrated* geospatial and temporal framework to provide basic information on the current condition of the state's natural resources for the purpose of socioeconomic development, maintenance of environmental stability and enhancement of the quality of life for the people of Jalisco. More specifically:

- Provide information on the extent and condition of forest resources with respect to forest cover, species distribution, volume, biomass and growth,
- Provide information on current land use patterns and assessment of the extent of environmental degradation,
- Provide an up-to-date land cover map, and
- Establish a mechanism to monitor the condition of the resources.

Due to the variety of needs and concerns of partner institutions involved in this project, as well as to the limited science and technology capacity in place to address the most relevant issues and problems, a pilot study covering over half a million hectares in central Jalisco was carried out to learn more about the different structural processes that are essential to the

design and implementation of a monitoring program for the assessment of terrestrial ecosystems. Clearly, the total experience and synthesis of this pilot study (Reich et al. 2005) were central to supporting IMRENAT's integrated sampling design and strategies for field data collection, quality assurance and quality control, management, and statistical analysis processes.

One of the objectives in designing IMRENAT was to employ a method capable of providing regional and local estimates of selected variables with the spectral and spatial precision of Landsat ETM+ imagery and with minimal input from any other system of characterizing the landscape. A two-way nested stratified sampling design was used to randomly allocate sample plots (30×30 m satellite pixel) throughout the state. The state was initially divided using two criteria, the first being the bioclimatic region which is highly correlated to the type of vegetation, and the second, if the sample plot, or pixel, was forested or non-forested. Separate Landsat ETM+ spectral classes were nested within each stratum. Previous studies have found significant relationships among spectral information with Landsat TM and ETM+ imagery with variables such as forest age, forest type, successional status, basal area, average tree height, density, volume and biomass (Brockhaus and Khorram 1992; Lee and Nakane 1996; Steininger 2000; Laurent et al. 2005). Studies have also shown that Landsat TM imagery provides comparable, and, in some cases, stronger predictions of certain forest structural features, such as basal area, when compared to radar satellite systems (Hyypä et al. 2000; Lefsky et al. 2001) or other optical sensors of similar spatial resolution (Brockhaus and Khorram 1992). If the sample captures this integrated multivariate variability this should allow one to apply these relationships to any geographical region provided information is available on the spatial and spectral variability of the area. IMRENAT is also capable of providing classical estimates of population means and variances. Because of the way IMRENAT was designed, it may not be as precise as a simple random sample of the same effort. It should be noted that one of the primary objectives in designing IMRENAT was the ability to monitor the heterogeneity (plants, animals, soils, species richness, pests, etc.) that exists within the state and not necessarily the most precise survey for estimating single population

parameters, say the total timber volume in the state. To the best of our knowledge, this application is the first of its kind that was designed and implemented to cover all land uses, including forests, agricultural lands, rangelands, urban lands, and standing water.

The purpose of this paper is to present the statistical design and estimation processes developed for the assessment of a variety of questions related to the condition and trends of important ecological resources in the State of Jalisco. Some initial results of the first phase of IMRENAT conducted in 2006 are presented and discussed. The data collected contains a wealth of information and is not possible to present all of the results in this paper. For a comprehensive presentation see FIPRODEFO (2007).

State of Jalisco

The State of Jalisco is located in west central Mexico and covers an area of approximately 7.9 million hectares (19.4 million acres). Four major ecological regions provide natural resource and environmental conditions that make this region one of the most prosperous in Mexico. Ecoregions include 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, are several important hydrological regions (watersheds) that drain to the Pacific Ocean (Lerma-Santiago, Huicicila, Ameca, Costa de Jalisco, Armeria-Coahuayana, Balsas, and El Salado). One of the watersheds, the Lerma-Santiago Hydrological Region, is connected to Chapala Lake, the most important source of water for the City of Guadalajara (INEGI 2001) and largest lake in Mexico.

The region's biophysical heterogeneity is reflected in a large diversity of plant and animal species. Within this region there are a significant number of species of mammals and birds, many of which are threatened by human activities. Some of the plant and animal species are endemic to specific locations. For example the pine-oak forests are home to "specialty" birds such as the thick-billed parrot, the Mexican-spotted owl, and a variety of woodpeckers. To a large extent Jalisco's economic growth and sustainability is anchored on the wealth and health of its natural resources; the monitoring and assessment of its

natural capital therefore is central to ecological and economic sustainability of the state at multiple geographical and organizational scales.

GIS and Landsat ETM+ data

Spectral information and topographic data were taken from a satellite imagery and Digital Elevation Model (DEM) of the state. Ten cloud-free Landsat-7 ETM+ images obtained during the months of January through March, 2004, were joined together to create a seamless image of the state. The image was normalized (Hall et al. 1991) to account for differences among scenes and used the normalized image for model development. The satellite imagery consisted of nine spectral bands [spectral bands 1–5, 6L (low gain), 6H (high gain, see USGS-EROS Data Center web site for more information), 7 and 8]. Spectral bands 6L and 6H were thermal bands (57 m resolution), while band 8 was a panchromatic image (15 m resolution). These latter three bands were resampled to a 30 m spatial resolution using nearest neighbor techniques (Muukkonena and Heiskanenb 2005). Nearest neighbor resampling was selected due to quicker computer processing time as compared to other interpolation methods. In addition, nearest neighbor interpolation better maintains original reflectance values while providing sufficient accuracy and reduced potential introduction of unwanted geometric distortions in areas with no ground control points to provide precise control (Muukkonena and Heiskanenb 2005).

The DEM was obtained from the National Elevation Dataset (NED) as a seamless ArcInfo (ESRI 1995) grid at a 90 m resolution (US Geological Survey (USGS), Gesch et al. 2002). The DEM was resampled to a 30 m spatial resolution using bilinear techniques (Edenius et al. 2003), producing a more continuous surface reflecting gradual changes in elevation at a 30 m spatial resolution.

Plot configuration

An important goal of IMRENAT was to provide spatial estimates for variables collected on the sample plots as well as to provide estimated means and totals for the population parameters. The ability to spatially interpolate a set of data depends on how it is collected

and should not be overlooked when designing an inventory and monitoring program. If one can design the survey to capture the spatial variability in the data by stratifying the sample space in an integrated multivariate way, this increases one's ability for interpolating the data. However, in well-established institutionalized inventory and monitoring programs, one does not have an opportunity to influence the data collection process and therefore resulting data may not be optimal in providing spatial estimates in spite of the fact that the data provides unbiased estimates of the population parameters. Technology and budget constraints often are serious barriers to the advancement of integrated sampling strategies and its implementation. Such is not the case in the current situation.

Data from multiple sources complicates the modeling process, especially if different sampling units are used. The size and shape of the sampling unit can also influence the ability to spatially interpolate a set of data. Because satellite imagery is commonly used to help interpolate field data, it is important to know if the imagery adequately describes the variability observed in the field data. If the spatial resolution of the field data does not correspond to the resolution of the imagery, this could create unwanted noise in the data and thus, make it more difficult to interpolate. The primary sampling unit was selected to correspond to the spatial resolution of satellite imagery (30×30 m). Each primary sampling unit consisted of nine 10×10 m secondary sampling units. This plot configuration has been used successfully in modeling forest stand structure on the Kaibab National Forest in northern Arizona (Reich et al. 2004b), and modeling forest fuel loadings on the Black Hills National Forest in South Dakota (Reich et al. 2004a). Each primary sampling unit was centered on the coordinates assigned to it (i.e., center of a Landsat EMT+ pixel) and was laid out in a north–south, east–west manner. With the aid of topographic maps, satellite imagery and GPS, field crews navigated to the appropriate locations with an estimated accuracy of within 3 m. This level of accuracy was required to minimize registration errors with the field plots and to insure that at least 87% of the sample plots fell within the designated pixel on the satellite imagery. Research by Dai and Khorram (1998) indicate that due to registration errors the accuracy of remotely sensed change detection can be substantially degraded. Results of their analysis on Landsat TM data

indicated that a registration error of less than one-fifth of a pixel (6 m) was required to achieve a change detection error of less than 10%. Registration errors of less than 3 m were deemed acceptable for the modeling process. All sample plots were documented on the ground. For a complete description of the sampling protocol see Martinez et al. (2007).

Essential features and key elements observed on the sample plots, with emphasis on the forest resources included:

- Diameter, height, cubic volume, biomass as well as features of the cover of all trees above a specified diameter to characterize the structure of the forests
- Diameter growth of individual trees in relation to location and conditions
- Topography
- Soils and geological substratum of the population, including the depth of the soils
- Ground vegetation, with special reference to all rare species, including grasses, shrubs, fungi, bryophytes, lichens, etc.
- Occurrences and extent of regeneration (seedlings and saplings)
- Wildlife indicators (bird calls, scat, tracts, nests, etc.)
- Nature and quantity of all fallen and standing dead trees and fallen woody material
- Human influence and the history of the population (fire, clearings, logging, etc.)

Sampling design

A two-way nested stratified design was used to sample the state. In some instances it may be natural or desirable to stratify a population to be sampled by two or more alternative criteria (Bryant et al. 1960, Cochran 1977, pp. 124–126). For instance, sample plots could be stratified by bioclimatic zone, or if the sample plots are forested or non-forested. The type and diversity of the vegetation is correlated with bioclimatic zones, while the type and amount of a resource can vary depending if it occurs on forest or non-forested lands. Both stratification criteria were used to create a $2C$ “two-way table of strata cells” (Bryant et al. 1960), where C is the number of bioclimatic zones. If the population fraction falling into the $2C$ strata cells are known, one can use these cells as strata (Bryant et al. 1960). In case of IMRENAT,

the number of sampling units, or pixels assigned to each strata cell is known. A two-way stratified design is an effective design if the population cell frequencies are proportional to the marginal frequencies. Under this scenario, the two-way stratification will generally be more precise than simple stratification using either of the two criteria for constructing a single set of strata (Bryant et al. 1960). This is not the case with IMRENAT. Also, if the sample is not proportionally allocated to the strata cells this may result in a loss of precision for certain characteristics by comparison with a completely random design.

To further account for the variability in the population, separate Landsat ETM+ spectral classes were established for each stratum, the nested component of the sampling design, resulting in a total of 2CS strata, where S is the number of spectral class in each of the 2C strata cells. This level of stratification is referred to in the literature as “deep stratification” (Tepping et al. 1943). The problem with using such a design is that some strata cells may not have any samples. Tepping et al. (1943) looked at the theoretical evaluations of the sampling errors and biases of several sampling designs involving deep stratification and compared them with some commonly used designs. The authors found that while deep stratification designs were substantially better than unrestricted random sampling, in some instances larger errors were observed indicating a loss of information over random sampling.

A description of the stratification process follows:

Bioclimatic zones Recently Reich et al. (2004c) identified 12 unique Bioclimatic Zones (BCZ) using climate (temperature, precipitation and evaporation), topography (elevation, slope and aspect) and vegetation. These zones coincide in general with those adopted to describe the vegetation and ecology in Mexico (Rzedowski 1978). The BCZs define three broad ecological regions (C=3). The first is the *sub-humid tropical region (SHTROPR)* which is located along the Pacific coast. SHTROPR is characterized by

hot temperatures, monsoon rains during the summer months and an annual dry period of 5–9 months. Tropical dry forests dominate this region (García 1965). At higher elevations is the *sub-humid temperate region (SHTEMPR)* which covers the greatest portion of the state. Pine-oak and mixed deciduous hardwood forests dominate this region. Going eastward away from the coast bioclimatic conditions change to an *arid and semi-arid environment (ASARIDR)*, which has a low annual precipitation with 6–8 dry months. Dominant vegetation includes mezquital–huizachal, matorral crasicaule and chaparral. The spatial variation in vegetation is consistent with the temperature and precipitation–evaporation zones used in defining the BCZs referred to above. The increase in tree species diversity (Table 1) from east to west is likely a result of increased precipitation (Gentry 1995). Rainfall and temperature have long been known as chief factors that determine tropical vegetation (Holdridge 1947).

The advantage of using bioclimatic zones to define strata is that they are ecologically important, well recognized and likely to remain constant over time. The BCZs also conform to some extent, to political boundaries within the state (e.g., regions and counties) making it simpler to allocate sample sizes in relation to local needs and available resources (Husser et al. 1997).

Forest or non-forest A binary regression tree (Breiman et al. 1984) was used to classify each pixel of the satellite imagery, or potential primary sampling unit in the state as being forested or non-forested with an overall accuracy of 91%. Forested areas were classified as areas dominated by woody vegetation. Non-forested areas included agricultural lands, bare soil, grasslands, urban areas, water, coastal vegetation, and orchards. Forested areas covered approximately 60% (±2%) of the state. By sub-dividing the BCZs into forest and non-forest strata allows for better monitoring of land-use conversions and in allocating sample sizes to reflect the importance of a region either socially or economically.

Table 1 Trees species diversity observed on forested plots in the network of permanent plots in Jalisco, Mexico by bioclimatic zone

Bioclimatic zone	Number of tree species counted on sample plots	Number of trees counted on sample plots	Diversity index (Shannon–Weaver Index)
Semi-Arid	88	1,185	3.40
Temperate	165	3,321	3.67
Tropical	214	2,963	4.48

Spectral variability To avoid predictive errors potentially fostered by inappropriate or inaccurate land cover maps, a logical approach to interpolate patterns in forest structure or any other key indicator variable was to use Landsat ETM+ data to establish separate spectral information for each stratum (Brockhaus and Khorram 1992; Lee and Nakane 1996; Hyypä et al. 2000; Steinger 2000). This level of stratification was intended only for the spatial statistical modeling component of the inventory and not necessarily to provide estimates for individual spectral classes. Spectral classes are used as a vehicle for defining domains for many types of sub-populations even if the boundaries of the sub-populations change over time (Husser et al. 1997). Spectral classes can be used to define, for example, land cover types (Joy et al. 2003), ecological niches (Binford et al. 2004), potential wildlife habitat areas (Reich et al. 2004b; Laurent et al. 2005), forest stand structure (Reich et al. 2004b) and forest fuel loadings (Reich et al. 2004a).

The spatial extent of each of the six 2-way strata were used to create a mask in ERDAS Imagine to subset the satellite imagery into six separate images, or strata. ERDAS was used to perform a maximum likelihood unsupervised classification on each of the images to identify $S=10$ Multivariate Spectral Classes (MSC) with similar spectral properties. All eight Landsat ETM+ bands were used as input. As a result of this analysis, each pixel, or potential sample plot was assigned to one of 60 strata.

Stratifying by bioclimatic zones and by forest or non-forest we found by experimentation that 10 spectral classes provided an optimum number of strata that could be effectively handled on an operational basis while capturing the spectral variability in the landscape (Binford et al. 2004). Defining more than 10 classes resulted in many classes containing a very small number of cells making it difficult to precisely locate in the field, while defining less than 10 classes did not adequately describe the variability of the vegetation and land cover types with a high degree of accuracy.

Sample size and allocation

Results from a pilot study (Reich et al. 2005) that encompassed 12 counties surrounding the City of Guadalajara, suggested that in forested areas, sample

sizes ranging from 300 to 600 were needed to obtain estimates of forest structure (basal area, cubic volume, biomass, etc.) with a reasonable bound on the error of estimation (15–20% at the 0.95 level of confidence). In allocating the sample plots, a target of at least 600 sample plots was set for the temperate bioclimatic zone because of its economic importance in relation to timber production and the fact that it covers the largest area. The tropical zone, the next most important area in terms of forest production was set a target of at least 400 sample plots, while the semi-arid region, which is the least important area in terms of forest production and economic importance, had a goal of at least 300 sample plots. In the non-forested regions, only categorical data was collected, such as the type of erosion, soil type, ground cover, etc., so a minimum sample size of 100 sample plots was set. Sample plots were randomly located within strata.

In 2005, 1,441 sample plots were established. Approximately 11% of the initial samples were inaccessible because of steep terrain, denied access by landowners, or potential danger to the field crews due to ongoing drug operations. The inaccessible plots may represent unique sub-populations and if information from these areas is not included in the survey, the statistics could be biased. To minimize this potential bias, sample plots were replaced by randomly selecting a new plot from a list of replacement plots that occur in the same bioclimatic zone, is forested or non-forested and is in the same spectral class as the plot being replaced. The replacement process is not perfect and probably does not entirely eliminate the potential of a bias, especially if the replacement plots are not truly representative of the inaccessible plots.

Roberts (2003) evaluated several imputation techniques for replacing missing field observations. These included ignoring the data, the use of previous observations, stratum means from observed sample plots, model predictions and nearest-neighbor imputations from observed plots. The author found that if the proportion of missing plots was less than 10% acceptable results were obtained for techniques that both ignored and replaced missing plot data, but only when the techniques accounted for differences between ownership categories. An alternative approach is to retain the inaccessible plots (this information is still part of the initial database) and use this information to form a new stratum (Cochran 1977,

pp. 359–360). This would allow for the estimation of the total area of inaccessible areas within the state and allow one to make assumptions about the characteristics of these areas. If the plots become accessible in the future, they can be easily be incorporated back into the survey. This aspect of the survey design needs to be investigated to address potential problems and its impact on the bias and precision of the stratified estimates.

Table 2 summarizes the final allocation of sample plots among the three bioclimatic zones, the forest/non-forest stratum and MSCs. Figure 1 depicts the location of the sample plots which shows the excellent geographic coverage obtained throughout the state.

Sample plots were located in a variety of forest types ($n=985$) including temperate forests {pine ($n=30$), oak ($n=211$), pine-oak ($n=111$), juniper-pine ($n=8$), fir ($n=2$), cloud forests ($n=3$)}, tropical forests {medium tropical ($n=131$), low tropical ($n=273$)}, subtropical shrublands ($n=121$), natural palms ($n=6$) and semi-arid forests {mezquital-huizachal ($n=79$)}.

Parameter estimation

In this section equations are presented for estimation of population parameters associated with IMRENAT. Before proceeding, some notation needs to be defined. Let $C=3$ represent the number of bioclimatic zones; $S=10$, the number of spectral classes (MSC), N_i = number of sampling units (pixels) in bioclimate zone i

($i=1, 2, 3$), $N = \sum_{i=1}^C N_i$ =number of sample units in the population (number of pixels), $W_i = N_i/N$ =stratum weights for the bioclimatic zones, $L=2$, for the forest/non-forest stratum, N_{ij} =number of sample units in forest/non-forest stratum j ($j=1,2$) in bioclimatic zone i , $W_{ij} = N_{ij}/N_i$ =stratum weights for the j th forest/non-forest stratum in bioclimatic zone i , n_{ij} =number of sample units in forest/non-forest stratum j in bioclimatic zone i .

To estimate the population parameters, post-stratification is used to allocate samples to predefined land-cover types which will be referred to as a domain. Domains differ from strata (i.e., spectral classes) in that the strata were used to allocate the samples in the field, while domains are the land cover types observed on the sample plots.

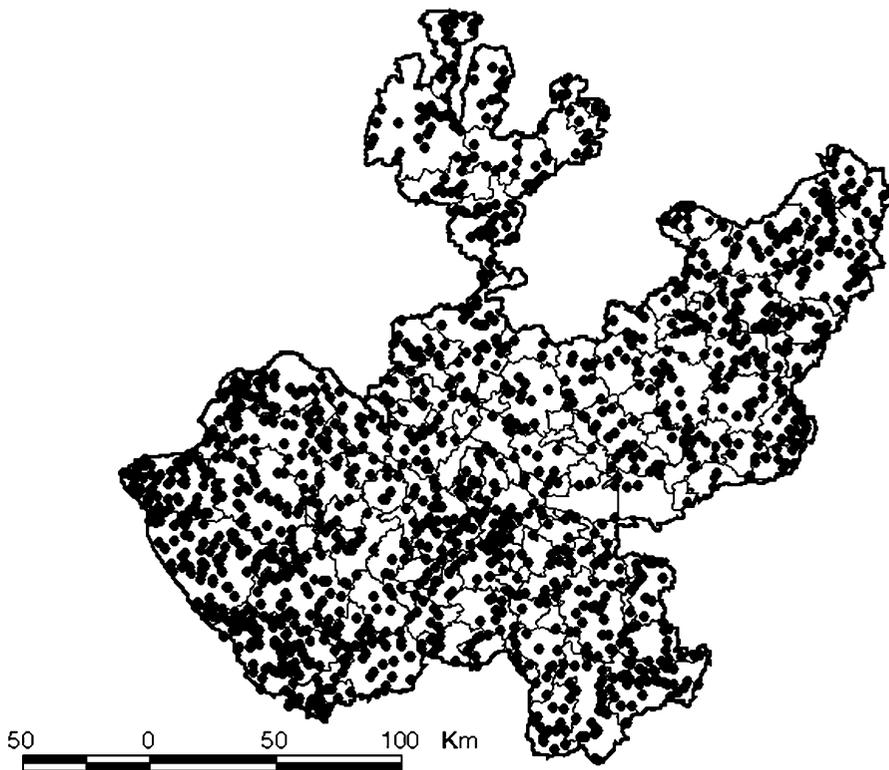
To simplify the notation, subscripts representing bioclimatic zones (i) and forest/non-forest stratum (j) will be temporarily suppressed. Thus, within a given bioclimate zone and forest/non-forest stratum, let $I_{dhk}=1$ if a sample unit k , in stratum h is in domain d , and 0 otherwise. Thus $n_{dh} = \sum_k I_{dhk}$ is the number of sample units in stratum h and domain d . Also, let n_h represent the number of sample units in stratum h , and n_d the number of sample units in domain d . An estimate of the population mean in domain d from stratum h has the general form:

$$\bar{y}_{dh} = \frac{\sum_{k=1}^{n_h} I_{dhk}y_{hk}}{n_{dh}} \tag{1}$$

Table 2 Distribution of sample plots in the network of permanent plots in Jalisco, Mexico by bioclimatic zone, the forest/non-forest stratum, and spectral class

Spectral class	Semi-arid		Temperate		Tropical	
	Non-forested	Forested	Non-forested	Forested	Non-forested	Forested
1	9	16	0	27	1	22
2	5	50	12	44	15	34
3	8	28	32	40	7	25
4	4	35	24	58	8	31
5	9	24	29	78	16	28
6	6	26	20	40	11	41
7	3	19	38	33	10	28
8	11	30	11	66	14	34
9	8	20	24	64	14	53
10	2	16	8	42	7	24
Total	65	264	198	492	103	320

Fig. 1 Spatial distribution of the network of permanent plots (n=1442) in Jalisco, Mexico



with sample variance s_{dh}^2 . The estimated population mean for domain d is given by (Cochran 1977, p 143, Eq. 5A.69)

$$\bar{y}_d = \frac{\sum_{h=1}^H \frac{N_h n_{dh}}{n_h} \bar{y}_{dh}}{\sum_{h=1}^H \frac{N_h n_{dh}}{n_h}} \tag{2}$$

with estimated variance (Cochran 1977, p 144, Eq. 5A.75):

$$\hat{V}(\bar{y}_d) = \frac{1}{N_d^2} \sum_{h=1}^H N_h^2 \left(\frac{(n_{dh} - 1)s_{dh}^2 + n_{dh} \left(1 - \frac{n_{dh}}{n_h}\right) (\bar{y}_{dh} - \bar{y}_d)^2}{n_d(n_d - 1)} \right) \tag{3}$$

where $N_d = \sum W_{ij}$ represents the number of sampling units in domain d in the i th bioclimate zone and j th forest/non-forest stratum. If the true stratum weights

(W_d) are unknown, one can use $w_d = \sum_{h=1}^H W_h \left(\frac{n_{dh}}{n_h}\right)$ as an estimate of W_d . Substituting,

$$\hat{V}(\bar{y}_d) = \left(\frac{1}{N_d^2 w_d^2} \right) \sum_{h=1}^H N_h^2 \left(\frac{(n_{dh} - 1)s_{dh}^2 + n_{dh} \left(1 - \frac{n_{dh}}{n_h}\right) (\bar{y}_{dh} - \bar{y}_d)^2}{n_d(n_d - 1)} \right) \tag{4}$$

Because of classification errors, it is possible for a given land cover type to occur in either the forest or non-forest stratum or both. Averaging across the forest and non-forest stratum within a given bioclimatic zone (i) we get

$$\bar{y}_{di} = \sum_{j=1}^L W_{ij} \bar{y}_{dij} \tag{5}$$

for an estimate of the population mean with variance

$$\hat{V}(\bar{y}_{di}) = \sum_{j=1}^L W_{ij}^2 \hat{V}(\bar{y}_{dij}) \tag{6}$$

Averaging across bioclimatic zones, an estimate of the population mean is given by

$$\bar{y}_{st} = \sum_{i=1}^C W_i \bar{y}_{di} \tag{7}$$

with estimated variance

$$\hat{V}(\bar{y}_{st}) = \sum_{i=1}^C W_i^2 \hat{V}(\bar{y}_{di}). \tag{8}$$

These equations are used to obtain parameter estimates at both, the stand level and individual tree level.

Table 3 summarizes selected statistics from IMRENAT. Estimates of average tree diameter (cm) and tree height (m) had percent sampling errors less than 10%, while estimates of trees/ha and basal area (m² ha⁻¹) had errors between 10 and 15%. The largest errors were associated with estimates of volume (m³ ha⁻¹) and biomass (tones/ha) (17% error). The percent sampling errors meet the goals of statistical precision set by IMRENAT of a 20% error at the 0.95 level of confidence. In general, the semi-arid region had the largest sampling error, followed by the tropical zone while the most precise estimates were obtained in the temperate zone. The level of precision was proportional to the sample sizes allocated to the individual bioclimatic zones.

Spatial statistical models

In recent years there has been an increasing interest in developing spatial statistical models for data sets that are seemingly spatially independent. This lack of spatial structure makes it difficult, if not impossible to use optimal predictors such as ordinary kriging for modeling the spatial variability in the data. In many instances, the data still contain a wealth of information that could be used to gain flexibility and precision in estimation. The inventory data was used to develop spatial statistical models of variables selected to describe forest structure using a combination of regression analysis to describe the large-scale spatial variability in a set of geodata and a tree-based stratification design to enhance the estimation process of the small-scale spatial variability (Reich et al. 2004b). With this approach, sample units (i.e., pixel of a satellite image) are classified with respect to predictions of error attributes into homogeneous classes, and the classes are then used as strata in the stratified analysis. Independent variables used as a basis of stratification included terrain data and satellite imagery. A decision rule was used to identify a tree size that minimized the error in estimating the variance of the mean response and prediction uncertainties at new spatial locations.

Let $z(s_i)$ represent a sample value of the target variable Z at spatial location s_i . Also, assume the sample data contains a set of auxiliary variables (predictors) X , the values of which are known for all units in the population. Multiple regression is used to

Table 3 Parameter estimates and percent sampling error for selected variables describing forest structure and environmental disturbances in Jalisco, Mexico

Statistic	Sample mean	Percent sampling error ^a
Percent forest lands	60.5	5.8
Average tree diameter (cm)	22.43	6.1
Average tree height (m)	9.09	7.8
Trees ha ⁻¹	169.06	12.8
Basal Area (m ² ha ⁻¹)	7.89	13.2
Cubic Volume (m ³ ha ⁻¹)	46.19	17.3
Biomass (tones ha ⁻¹)	58.18	17.1
Percent of the state with no soil disturbances	34.1	12.1
Percent of the state with no site disturbances	28.8	13.6

^a The percent sampling error assumes a 0.95 level of confidence

describe the large-scale spatial variability in the data as a linear function in p known explanatory variables $x_j(s_i)$

$$z(s_i) = \beta_0 + \sum_{j=1}^p x_j(s_i)\beta_j + \eta(s_i) \quad (9)$$

where β_j , $j=0, \dots, p$ are $p+1$ unknown regression coefficients and $\eta(s_i)$ is an error process sometimes referred to as a random field, with $E[\eta(s_i)] = 0$ and covariance $C(x, y) = \text{Cov}(\eta(x_i), \eta(y_j))$. The error term in Eq. 1 is unknown because the true model is unknown. Once the model parameters have been estimated, the regression residuals are defined as $\hat{\eta}(s_i) = z(s_i) - \hat{z}(s_i)$, where $\hat{z}(s_i)$ is the predicted value at spatial location s_i given the explanatory variables $x_j(s_i)$. The error process can be expressed as (Cocchi et al. 2002):

$$\eta(s_i) = \hat{\eta}(s_i) + \mu(s_i) \quad (10)$$

with $E[\mu(s_i)|\hat{\eta}(s_i)] = 0$. Using the set of auxiliary variables, X as a basis of stratification assume

$$\hat{\eta}(s_i) = f(x(s_i)) + \delta(s_i) \quad (11)$$

with $E[\delta(s_i)|x(s_i)] = 0$, $f(x(s_i))$ is a deterministic function, and $\delta(s_i)$ is a zero-mean random term (Cocchi et al. 2002). Combining Eqs. 10 and 11

$$\eta(s_i) = f(x(s_i)) + \varepsilon(s_i) \quad (12)$$

with $E[\eta(s_i)|f(x(s_i))] = f(x(s_i))$, $\varepsilon(s_i) = \mu(s_i) + \delta(s_i)$, and $E[\varepsilon(s_i)|x(s_i)] = 0$ provided that $\mu(s_i)$ and $\delta(s_i)$ are conditionally independent (Cocchi et al. 2002; Benedetti et al. 2005). The mean function $f(x(s_i))$ is estimated by \hat{f} using the recursive partitioning method introduced by Brieman et al. (1984). Combining Eqs. 9 and 12 the full model describing the spatial variability in the sample data is given by

$$z(s_i) = \beta_0 + \sum_{j=1}^p x_j(s_i)\beta_j + f(x(s_i)) + \varepsilon(s_i). \quad (13)$$

Maps representing the components of forest structure were generated for the models selected to minimize the error in estimating the uncertainty in the spatial estimates. Maps displaying the uncertainty in the spatial estimates were also generated. State-wide estimates of the mean response and prediction variances were obtained by averaging estimates over all 30×30 m pixels in the state ($N=95,693,043$). The spatial estimates are self weighting and have the same weights as the classical estimates. These estimates were compared to classical estimates obtained using the same set of data.

Table 4 compares classical estimates for canopy closure (%), basal area ($\text{m}^2 \text{ha}^{-1}$), cubic volume ($\text{m}^3 \text{ha}^{-1}$), and biomass (tonnes/ha) with estimates obtained from the spatial statistical models. Figure 2 depicts an example of the spatial model developed for

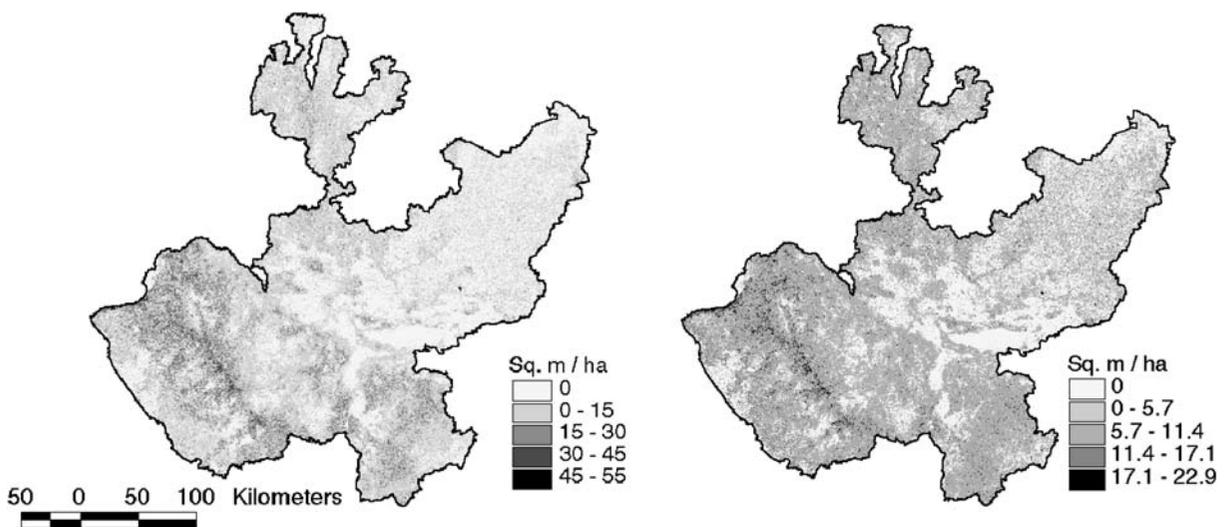


Fig. 2 Spatial distribution of predicted basal area (m^2/ha) (left) and associated prediction standard deviations (right) in Jalisco Mexico. The spatial statistical models accounted for 62% of the observed variability in basal area

Table 4 Comparison of a classical approach and a spatial statistical approach for estimating population parameters of selected variables describing forest structure in Jalisco, Mexico using data from IMRENAT

Variable	Classical estimation approach					Spatial estimation approach					
	\bar{y}^a	$s_{\bar{y}}^b$	$Bs_{\bar{y}}^c$	Lower bound ^d	Upper bound ^d	\bar{y}_{new}^e	Discrepancy ratio ^f	Min \hat{y}_{new}^g	Max \hat{y}_{new}^g	Std. dev. ^g \hat{y}_{new}^g	Std. error ($n=985$)
Canopy closure (%)	48.9	2.28	4.47	44.4	53.4	52.2	0.75	0.0	100.0	22.59	0.76
Basal area (m ² ha ⁻¹)	7.89	0.53	1.04	6.85	8.93	8.05	0.87	0.00	51.80	7.43	0.61
Cubic volume (m ³ ha ⁻¹)	46.2	4.08	8.00	38.2	54.2	47.2	0.12	0.0	526.0	66.24	0.56
Biomass (tones ha ⁻¹)	55.18	5.09	9.98	45.20	65.16	56.10	0.09	0.00	882.94	67.47	0.35

^a Estimated mean of the state-wide inventory using the classical estimation approach

^b Estimated standard error of the mean for the state-wide inventory using the classical estimation approach

^c $B=1.96$ for a 0.95 confidence bound

^d Lower and upper confidence bounds for the state-wide estimate

^e Estimated state-wide mean using the spatial estimation approach

^f Discrepancy ratio which measures the half width difference between the classical estimate and the spatial estimate. A value greater than 1 indicates the estimates from the spatial model are significantly larger than reported by the state-wide inventory, while a value less than -1 indicates estimates from the spatial statistical model are significantly less than reported by the state-wide inventory

^g Standard deviation of estimates from the spatial statistical models. This is not a measure of the precision of the spatial estimates

Estimated standard error of the mean for the spatial statistical models ($n=985$). The estimated standard errors are comparable to estimated standard errors reported for the state-wide inventory using the classical approach

basal area. Basal areas ranged from 0 to 52 m² ha⁻¹ with the largest basal areas occurring at the higher elevations and in the tropical forests in the northwestern part of the state (Fig. 2).

Except for one case, no significant differences ($p > 0.05$) were observed between the classical estimates and the spatial estimates at the state level or by bioclimatic zones (not shown). The only significant difference was observed in canopy closure in the tropical bioclimatic zone. The spatial model predicted a denser canopy when compared to the classical estimates. At the regional level, numerous strata had missing data, especially in the semi-arid region. To obtain classical estimates, missing data were replaced using sample means from similar stratum in neighboring regions. When the classical estimates were compared to the spatial estimates (not shown), there was no significant difference for all variables in the temperate bioclimatic zone. The classical estimates suggest the spatial models under-estimated in some regions in the semi-arid bioclimatic zone and over-estimated in some regions in the tropical bioclimatic zone.

While the classical estimates are unbiased at the state level and bioclimatic region, the quality of information obtained for the regional estimates are unknown because of the method used in replacing missing values. The discrepancy between the classical estimates and the spatial models at the regional level suggest the need to evaluate the methods used for imputing regional estimates in the semi-arid and tropical bioclimatic zones. Since estimates are desired at the regional and county level in which few sample plots are available, some form of imputation (Schreuder et al. 1993) or small area estimation (Moisen et al. 2005) is required. In theory, the regression or hotdeck procedures should provide for all the needs of IMRENAT’s imputing needs. A set of rules would need to be developed to minimize the subjectivity required on the part of the user when imputing values (Van Duesen 1997). It is anticipated that these rules would evolve over time as details are refined and alternative methods evaluated. An overview of imputing methods can be found in Schreuder et al. (1993) and Van Duesen (1997).

Instead of imputing the missing data one could use small area estimation (Moisen et al. 2005) to obtain

estimates at the regional or county level. Moisen et al. (2005) describes how maps of forest/non-forest and forest biomass developed from satellite imagery when combined with field data can be used to produce estimates of forest biomass for small geographical regions (i.e., regions or counties). The process involves identifying an area of interest and then construct estimates within these areas by creating weighted combinations of direct estimates using the field data and indirect estimates using for example, estimates from the spatial statistical models described above, using a composite estimator (Cochran 1977, p.352). The composite estimator weights the estimates inversely proportional to their variance. In the case of the spatial statistical models variance estimates can be obtained directly from the error surfaces (Table 2, Fig. 2), while the variance estimates for the direct estimates can be obtained using the field data. The feasibility of this approach still needs to be evaluated.

The agreement between the classical estimates and the spatial estimates at the regional, bioclimatic and state level suggest that it might be possible to use the spatial statistical models to provide reasonable estimates of population parameters even at the county level. This was one of the primary goals of IMRENAT; the ability to spatially model the distribution of selected variables and provide estimates for any geographical region within the state. The spatial extent to which the spatial statistical models can provide reliable estimates still needs to be evaluated. Surfaces of the prediction standard deviations (Fig. 2) are being developed to evaluate the predictive performance of the models as well as identifying regions in which additional samples should be allocated to increase the precision of the estimates.

The future

The bioclimatic and multi-spectral-based sampling strategy that has been implemented in Jalisco is a considerable advance over historical sampling strategies implemented in other countries. It provides an adaptive approach for interrelating designed-based and model-assisted estimation processes into one integrated sampling strategy. Both classical and spatial estimates can be generated due to the integrative and multivariate nature of the sampling strategy. The monitoring aspect

of IMRENAT will also allow one to model the spatiotemporal variability in selected key variables.

Spatiotemporal data and information that is multivariate in their structural nature is critical for decision making by state officials and industries. Such information has never been available in other surveys of this type with any degree of reliability and consistency. For example, researchers and ecosystem resource managers can benefit greatly from the utilization of integrative sampling-based geoinformation due to the wealth of spatiotemporal dependencies that are available for studying and modeling complex ecological relationships. The ability to keep track of these ecological and socioeconomic complexes will require large capacity and highly efficient computerized geographic information systems. These geographical information systems will also have to be integrated with equally complex software for analysis of remotely sensed and field measurement data.

Since IMRENAT is a brand new approach one can expect changes to be made over time to improve conceptual and operational design aspects. We will investigate various new remote sensing capabilities to further improve getting detailed local information, more sophisticated analysis to satisfy the requirements of officials as they become more aware of the capabilities of the system, possible improvements in change assessments over time, and opportunities to focus on specific interesting changes that may or may not be occurring in Jalisco when the sample size is too small to be sure of what is going on in a particular part of the state or with a particular resource.

In light of its integrative nature, there ought to be considerable effort put into training personnel, developing strong data collection and analysis capabilities, and attracting good researchers to work with the data and data collection approach. As this study indicates, current and future information technology advancements provide a solid prospect for the development and application of integrative sampling strategies to support sustainability science and management processes.

Acknowledgements The authors wish to thank the state government of Jalisco (SEDER, DGFS, and FIPRODEFO) for funding the IMRENAT program and supporting the generation of these results. We also thank those CAMESA-related federal agencies of the NAFTA countries (USDA FS, USEPA, NRC CFS, and EC-EMAN) and scientists for their participation in

supporting this effort. The authors would also like to acknowledge the many valuable suggestions from the anonymous reviewers.

References

Benedetti, R., Espa, G., & Lafratta, G. (2005). A tree-based approach to forming strata in multipurpose business surveys. Discussion Paper No. 5 2005, Dipartimento di Economia, Università Degli Studi di Trento, Trento, Italy. 17 pp.

Binford, M. W., Lee, T. E., & Townsend, R. M. (2004). Sampling design for an integrated socioeconomic and ecological survey by using satellite remote sensing and ordination. *Proceedings of the National Academy of Science*, *101*, 11517–11522.

Brieman, L., Freidman, J., Olshen, R., & Stone, C. (1984). *Classification and regression trees*. Pacific Grove, CA: Wadsworth and Brooks 358 p.

Brockhaus, J. A., & Khorram, S. (1992). A comparison of spot and Landsat-TM data for use in conducting inventories of forest resources. *International Journal of Remote Sensing*, *13*, 3035–3043.

Bryant, E. C., Hartley, H. O., & Jessen, R. J. (1960). Design and estimation in two-way stratification. *Journal of the American Statistical Association*, *55*, 105–124.

Cocchi, D., Fabrizi, E., Raggi, M., & Trivisano, C. (2002). Regression trees based stratification: An application to the analysis of the Italian post enumeration survey. Proceedings of the International Conference on Improving Surveys, August 25–28, 2000, Copenhagen, Denmark. <http://www.icis.dk/ICIS-papers/B2_5_2.pdf>

Cochran, W. G. (1977). *Sampling techniques*. New York: Wiley 428 pp.

Dai, X., & Khorram, S. (1998). The effects of image misregistration on the accuracy of remotely sensed change detection. *IEEE Transactions on Geoscience and Remote Sensing*, *36*, 1566–1577.

Edenius, L., Vencatasawmy, C. P., Sandstrom, P., & Dahlberg, U. (2003). Combining satellite imagery and ancillary data to map snowbed vegetation important to Reindeer *Rangifer tarandus*. *Arctic, Antarctic and Alpine Research*, *35*, 150–157.

ESRI (1995). *ARC/INFO® Software and on-line help manual*. Redlands, CA: Environmental Research Institute, Inc.

FIPRODEFO (2007). Reporte Jalisco 2006. Fideicomiso para la Administración del Programa de Desarrollo Forestal, Gobierno del Estado de Jalisco. CD.

García, E. (1965). Distribución de la precipitación en la República Mexicana. *Publicaciones del Instituto de Geografía U. N. A. M.*, *1*, 173–191.

Gentry, A. H. (1995). Diversity and floristic composition of neotropical dry forests. In S. H. Bullock, H. A. Mooney, & E. Medina (Eds.) *Seasonally dry forests* (pp. 146–194). Cambridge, UK: Cambridge University Press.

Gesch, D., Oimoen, M., Greenlee, S., Nelson, C., Steuck, M., & Tyler, D. (2002). The national elevation dataset. *Photogrammetric Engineering & Remote Sensing*, *68*, 5–32.

Hall, F. G., Strelbel, D. E., Nickeson, J. E., & Gotz, S. J. (1991). Radiometric reflection: Toward a common radiometric

response among multidata multisensor images. *Remote Sensing of Environment*, *35*, 11–27.

Holdridge, L. R. (1947). Determination of the world plant formations from simple climatic data. *Science*, *105*, 267–368.

Husser, S. M., Breidt, F. J., & Fuller, W. A. (1997). Design and estimation for investigating the dynamics of natural resources. *Ecological Application*, *8*, 234–245.

Hyypä, J., Hyypä, M., Inkinen, M., Engdahl, M., Linko, S., & Zhu, Y. H. (2000). Accuracy comparison of various remote sensing data sources in the retrieval of forest stand attributes. *Forest Ecology and Management*, *128*, 109–120.

INEGI (2001). Estudio hidrológico del estado de Jalisco. Instituto Nacional de Estadística Geografía e Informática (INEGI), Segunda Edición, México, D.F., México.

Joy, S. M., Reich, R. M., & Reynolds, R. T. (2003). A non-parametric supervised classification of vegetation types on the Kaibab National Forest using decision trees. *International Journal of Remote Sensing*, *24*, 1835–1852.

Laurent, E. J., Shi, H., Gatzolis, D., Lebouton, J. P., Walters, M. B., & Liu, J. (2005). Using the spatial and spectral precision of satellite imagery to predict wildlife occurrence patterns. *Remote Sensing of Environment*, *97*, 249–262.

Lee, N. J., & Nakane, K. (1996). Forest vegetation classification and biomass estimation based on Landsat TM data in a mountain region of west Japan. In H. L. Gholz, K. Nakane, & H. Shimoda (Eds.) *The use of remote sensing in the modeling of forest productivity* (pp. 159–171). Dordrecht, The Netherlands: Kluwer.

Lefsky, M. A., Cohen, W. B., & Spies, T. A. (2001). An evaluation of alternate remote sensing products for forest inventory, monitoring and mapping of Douglas-fir forests in western Oregon. *Canadian Journal of Forest Research*, *31*, 78–87.

Martínez, A., Flores, G., Talavera, E., Alonso, A., Flores, L. J., & Aguirre-Bravo, C. (2007). Manual para la toma de datos de campo: Proyecto de Inventario y Monitoreo de los Recursos Naturales de Jalisco. Version 3.0. Fideicomiso para la Administración del Programa de Desarrollo Forestal del Estado de Jalisco (FIPRODEFO), Guadalajara Jalisco, Mexico. 138p.

Moisen, G. G. Blackard, J. A., & Finco, M. (2005). Small area estimation in forests affected by wildfire in the Interior West. In J. D. Greer (Ed.), *Remote sensing for field users: Proceeding of the Tenth Forest Service Remote Sensing Applications Conference*; 2004 April 5–9, Salt Lake City, UT. Bethesda, MD; American Society of Photogrammetry and Remote Sensing. CD-ROM. USDA Forest Service – Research & Development.

Muukkonena, P., & Heiskanen, J. (2005). Estimating biomass for boreal forests using ASTER satellite data combined with standwise inventory data. *Remote Sensing of Environment*, *99*, 434–447.

Reich, R. M., Aguirre-Bravo, C., & Bravo, V. A. (2004c). Climate modeling for the states of Jalisco and Colima, Mexico. Fideicomiso para la Administración del Programa de Desarrollo Forestal del Estado de Jalisco, Guadalajara, Jalisco, Mexico, 115 pp.

Reich, R. M., Aguirre-Bravo, C., Bravo, V. A., Mendoza, M. A. B., Garnica, G. F., Rodríguez, L. J. F., et al. (2005). Análisis de estrategias estadísticas para el avance del Monitoreo y

- dictaminación de los ecosistemas terrestres en el estado de Jalisco, Mexico: Síntesis del estudio piloto. Documento Técnico 43. FIPRODEFO, Guadalajara Jalisco, Mexico. 99p CD.
- Reich, R. M., Joy, S. M., & Reynolds, R. T. (2004b). Predicting the location of northern goshawk nests: Modeling the spatial dependency between nest locations and forest structure. *Ecological Modeling*, *176*, 109–113.
- Reich, R. M., Lundquist, J. E., & Bravo, V. A. (2004a). Spatial models for estimating fuel loads in the Black Hills, South Dakota, USA. *International Journal of Wildland Fire*, *13*, 119–129.
- Roberts, R. E. (2003). Compensating for missing plot observations in forest inventory estimation. *Canadian Journal of Forest Research*, *33*, 1990–1997.
- Rzedowski, J. (1978). *Vegetación de México*. Editorial Limusa. México, D.F, Mexico.
- Schreuder, H. T., Gregoire, T. G., & Wood, G. B. (1993). *Sampling methods for multiresource forest inventory*. New York: Wiley 446p.
- Steininger, M. K. (2000). Satellite estimation of tropical secondary forest above-ground biomass: Data from Brazil and Bolivia. *International Journal of Remote Sensing*, *21*, 1139–1157.
- Tepping, B. J., Hurwitz, W. N., & Deming, W. E. (1943). On the efficiency of deep stratification in block sampling. *Journal of the American Statistical Association*, *38*, 93–100.
- Van Deusen, P. C. (1997). Annual forest inventory statistical concepts with emphasis on multiple imputations. *Canadian Journal of Forest Research*, *27*, 379–384.