

COMPARACIÓN DE ESTIMADORES PARA VOLUMEN MADERABLE EN SELVA MEDIANA DEL SURESTE DE MÉXICO

COMPARISON OF WOOD VOLUME ESTIMATORS IN MEDIUM HEIGHT TROPICAL FORESTS OF MEXICO

Efraín Velasco-Bautista¹, Héctor de los Santos-Posadas^{1*}, Hugo Ramírez-Maldonado², Gilberto Rendón-Sánchez¹

¹Campus Montecillo. Colegio de Postgraduados. 56230. Montecillo, Estado de México (velasco.efrain@colpos.mx) (hectorm.delossantos@gmail.com). ²DICIFO-Universidad Autónoma Chapingo. 56230. Chapingo, Estado de México.

RESUMEN

En México, el proyecto Inventario Nacional Forestal y de Suelos (INFyS) utiliza para el análisis estadístico de los datos dasométricos estimadores de razón cuya aplicación es fácil. Sin embargo, en circunstancias diversas podría ser imposible obtener datos que cumplan por completo con sus supuestos teóricos. En este estudio se compararon los resultados del estimador de razón del INFyS con estimadores alternativos. Con datos del 2010 provenientes de la malla de muestreo obtenida en selva mediana subperennifolia y subcaducifolia en Quintana Roo y Campeche, México, se compararon estadísticamente los estimadores: razón de medias, medias de razones e Inventario Forestal y Análisis (FIA-USDA). El sesgo relativo del estimador de razón (sesgo respecto al error estándar) en todos los casos fue menor a 10 %; por tanto se consideró insignificante. En las condiciones estudiadas, los resultados indican que los tres estimadores no son estadísticamente diferentes. Sin embargo, el estimador medias de razón presenta la ventaja de proporcionar estimaciones por hectárea de manera directa. Además, su planteamiento teórico permite analizarlo desde una perspectiva del estimador Horvitz-Thompson.

Palabras clave: Razón de medias, media de razones, bootstrap, zonas de inclusión espacialmente disjuntas.

INTRODUCCIÓN

En el proyecto del Inventario Nacional Forestal y de Suelos (INFyS) se utiliza el diseño de muestreo estratificado sistemático por conglomerados. Éstos se localizan en equidistancias

ABSTRACT

In Mexico, the project National Forest and Soil Inventory (Inventario Nacional Forestal y de Suelos-INFyS) uses ratio estimators that are easily applicable for statistical analysis of dasometric data. However, under diverse circumstances, it may be impossible to obtain data that fully comply with their theoretical assumptions. In this study, the results of the INFyS ratio estimator are compared with results using alternative estimators. With 2010 data from the sampling grid obtained in the medium height sub-evergreen and sub-deciduous forests of Quintana Roo and Campeche, Mexico, the ratio of means estimator, mean of ratios estimator and the Forest Inventory and Analysis (FIA-USDA) were compared statistically. The relative bias of the ratio estimator (bias, relative to the standard error) in all cases was below 10 % and, therefore, considered insignificant. Under the conditions of this study, the results indicate that the three estimators are not statistically different. However, the mean of ratios estimator has the advantage of directly providing estimations per hectare. Moreover, its theoretical approach permits its analysis from a perspective of the Horvitz-Thompson estimator.

Key words: Ratio of means, mean of ratios, bootstrap, spatially disjointed inclusion zones.

INTRODUCTION

The National Forest and Soils Inventory (Inventario Nacional Forestal y de Suelos-INFyS) uses the systematic stratified cluster sampling design. The clusters are located 5 km equidistance from each other in temperate forests, as well as in high and medium high tropical forests, 10 km in low tropical forests and semiarid communities, and 20 km in arid communities (CONAFOR, 2012).

*Autor responsable ❖ Author for correspondence.

Recibido: marzo, 2015. Aprobado: septiembre, 2015.

Publicado como ARTÍCULO en *Agrociencia* 50: 119-132. 2016.

de 5 km en bosques, así como en selvas altas y medianas, de 10 km en selvas bajas y comunidades semiáridas, y de 20 km en comunidades áridas (CONAFOR, 2012).

El conglomerado o unidad de muestreo primaria (UMP), desde el punto de vista conceptual, es una parcela circular de 1 ha (56.42 m de radio), en la cual se evalúan cuatro unidades de muestreo secundarias (UMS) o sitios, de 400 m² cada una, dispuestas geoméricamente en forma de una Y invertida con respecto al norte; éstas son circulares en el caso de bosques y vegetación de zonas áridas, y rectangulares en selvas. La UMS número 1 es el centro de la UMP y el punto de georeferenciación del conglomerado; las UMS 2, 3 y 4 son periféricas (satélite). La distancia del centro de la UMS 1 (sitio 1) al de las otras UMS (sitio 2, 3 y 4) es 45.14 m cuando los sitios son circulares y 36.42 m cuando son rectangulares. El azimut para localizar las UMS 2, 3 y 4 a partir del centro de la UMS 1 es 0°, 120° y 240°, respectivamente. Las unidades de estudio (UE) son los árboles u otros elementos biológicos en las UMS (Velasco *et al.*, 2005; CONAFOR, 2012). En cada UMS se mide el diámetro normal (DN) y la altura total de los árboles mayores a 7.5 cm de DN. El diseño del conglomerado es similar al del programa de Inventario Forestal y Análisis (FIA) para la evaluación de recursos forestales en EE.UU. (Bechtold y Scott, 2005; McRoberts *et al.*, 2005).

La ubicación de los conglomerados en campo está sujeta a las condiciones y la vegetación del terreno. En un conglomerado podría ser imposible instalar alguna o varias de las UMS, ya que se ubicarían en barrancas, lagos, terrenos agrícolas, caminos o minas (UMS inexistentes). En otras circunstancias, por escala en la cartografía base (Uso de Suelo y Vegetación SERIE IV de INEGI, 1:250 000) es posible que en campo una parte de conglomerado se ubique en un estrato diferente al definido en la cartografía. En otros casos, debido a los linderos o las barreras topográficas solo se puede instalar una parte del conglomerado y entonces las UMS serán inaccesibles. Esta situación de conglomerados de tamaño diferente ha conducido al uso de estimadores de razón en el INFyS.

Aun así cuando un conglomerado comprende dos o más subparcelas disjuntas es tratado como una sola parcela debido a que el conglomerado es atado a un solo punto muestral. En efecto, el conglomerado de

The cluster, or primary sampling unit (PSU), from a conceptual point of view, is a 1 ha circular plot (56.42 m radius), in which four secondary sampling units (SSU), or sites, are evaluated. SSU are 400 m² each, placed geometrically in an inverted Y-shape with respect to the north. They are circular in the case of temperate forests and arid zone vegetation, and rectangular in tropical forests. SSU number 1 is the center of the PSU and is the georeferencing point of the cluster; SSU 2, 3 and 4 are peripheral (satellite). The distance from the center of SSU 1 (site 1) to the center of the other SSU (sites 2, 3 and 4) is 45.14 m when the sites are circular and 36.42 m when they are rectangular. The azimuth for locating SSU 2, 3 and 4 from the center of SSU 1 is 0°, 120° and 240°, respectively. The study units (SU) are the trees or other biological elements in the SSU (Velasco *et al.*, 2005; CONAFOR, 2012). In each SSU, diameter at breast height (DBH) and total height of trees with DBH larger than 7.5 cm are measured. The cluster design is similar to the Forest Inventory and Analysis (FIA) program for evaluation of forest resources in the USA (Bechtold and Scott, 2005; McRoberts *et al.*, 2005).

Location of the clusters in the field is subject to conditions and vegetation of the terrain. In a cluster, it might be impossible to install one or several SSU since they might be placed in canyons, lakes, farmland, roads or mines (non-existent SSU). In other circumstances, by scale in the cartographic base (Land use and Vegetation SERIES IV, INEGI, 1:250 000), it is possible that in the field part of the cluster is located in a stratum different from that defined in the cartography. In other cases, because of borders or topographic barriers, only a part of the cluster can be installed, and thus, the SSU will be inaccessible. This situation of clusters of different sizes has led to the use of ratio estimators in the INFyS.

Even so, when a cluster comprises two or more disjointed subplots, it is treated as a single plot because the cluster is tied to a single sampling point. Indeed, the cluster of plots serves to identify those elements whose spatially disjointed inclusion zones include the sample point (Gregoire and Valentine, 2008). Thus, there are different ways to estimate the forest parameters of interest.

The contribution of medium height rainforest is poor to the reality and potential of lumber production, although it is useful to obtain some non-

parcelas sirve para identificar aquellos elementos cuyas zonas de inclusión espacialmente disjuntas incluyen al punto muestral (Gregoire y Valentine, 2008). Así, hay formas diferentes para estimar los parámetros de interés forestal.

La selva mediana aporta poco a la realidad y potencial de producción maderable, pero es útil para obtener algunos productos forestales no maderables. Además de la biodiversidad, la selva mediana subperennifolia es de importancia mínima en cuanto a su capacidad de almacenamiento de carbono. Eaton y Lawrence (2009) reportan que en la selva mediana subperennifolia de Campeche y Quintana Roo el carbono en la biomasa viva sobre el suelo oscila de 4.8 Mg ha^{-1} en un bosque secundario de tres años hasta 73.5 Mg ha^{-1} en un bosque maduro, mientras que el carbono aéreo combinado con el carbono del suelo varía de 192 Mg ha^{-1} en un bosque secundario de 12 años hasta 469 Mg ha^{-1} en selva madura. Estas cualidades reducen la importancia de estudiar a la selva mediana subperennifolia desde el punto de vista cuantitativo.

Aunque la relación lineal entre la variable de interés (volumen, área basal, número de árboles) y la auxiliar (área) es débil y la recta no pasa siempre por el origen, el estimador de razón se usa en el proyecto de Inventario Nacional Forestal y de Suelos (INFyS) de México. Por lo tanto y para investigar el comportamiento de estimadores alternativos en ecosistemas forestales complejos, el objetivo del presente estudio fue comparar los estimadores de medias de razones, razón de medias, y FIA, y el parámetro de interés fue el volumen maderable ($\text{m}^3 \text{ ha}^{-1}$) en selva mediana subperennifolia y subcaducifolia de Quintana Roo y Campeche, México. La hipótesis fue que los tres estimadores no son estadísticamente diferentes.

MATERIALES Y MÉTODOS

Diseño del conglomerado y zona de inclusión

El diseño del conglomerado del INFyS es un conglomerado de subparcelas (circulares o rectangulares) que comprende una subparcela central georeferida en un punto muestral (x_s, z_s) y $n-1$ subparcelas satélite no traslapadas en un arreglo fijo. Este tipo de arreglo de subparcelas define una zona de inclusión espacialmente disjunta alrededor del punto central de cualquier árbol (Valentine *et al.*, 2006). Un conglomerado de subparcelas incluye un

timber forest products. Besides biodiversity, the sub-evergreen medium height tropical forest is negligible in terms of its capacity for carbon storage. Eaton and Lawrence (2009) report that in the medium height sub-evergreen tropical forest of Campeche and Quintana Roo live biomass on the soil oscillates between 4.8 Mg ha^{-1} in a three-year-old secondary forest and 73.5 Mg ha^{-1} in a mature forest, whereas, aboveground carbon combined with soil carbon varies from 192 Mg ha^{-1} in a 12-year-old secondary forest to 469 Mg ha^{-1} in a mature rain forest. These qualities diminish the importance of studying the medium height sub-evergreen tropical forest from a quantitative perspective.

Although the linear relationship between the variable of interest (volume, basal area, number of trees) and the auxiliary (area) is weak and the line does not always pass through the intercept, the ratio estimator is used in the National Forest and Soils Inventory (INFyS) project of Mexico. Therefore, and to evaluate the behavior of alternative estimators in complex forest ecosystems, the objective this study was to compare the estimators means of ratios, ratio of means and FIA, and timber volume ($\text{m}^3 \text{ ha}^{-1}$) was the parameter of interest in sub-evergreen and sub-deciduous tropical forests of Quintana Roo and Campeche, Mexico. The hypothesis was that the three estimators are not statistically different.

MATERIALS AND METHODS

Cluster design and inclusion zone

The INFyS cluster design is a cluster of subplots (circular or rectangular) that comprises a central georeferenced subplot at a sampling point (x_s, z_s) and $n-1$ satellite non-overlapping subplots in a fixed arrangement. This type of arrangement of subplots defines an inclusion zone, spatially disjointed around the central point of any tree (Valentine *et al.*, 2006). A cluster of subplots includes one tree if the sampling point falls within any part of the inclusion zone of the tree (Valentine *et al.*, 2006; Gregoire and Valentine, 2008).

In order to present the possible estimators and their respective variances, in the following section presents the notation in detail.

Notation

A : forested area of the region of interest (ha); U_k : k^{th} element in the population; m : sampling points; P_s : s^{th} cluster, $s=1,2,\dots, m$;

árbol si el punto muestral cae en cualquier parte de la zona de inclusión del árbol (Valentine *et al.*, 2006; Gregoire y Valentine, 2008).

Con el fin de presentar los posibles estimadores y sus respectivas varianzas enseguida se detalla la notación.

Notación

A : área forestal de la región de interés (ha); U_k : k -ésimo elemento en la población; m : puntos de muestreo; P_s : s -ésimo conglomerado, $s=1,2,\dots, m$; (x_s, z_s) : coordenadas del conglomerado P_s ; (x_k, z_k) : ubicación geográfica de un elemento de la población, U_k ; a_{sj} : área de cada subparcela (400 m^2) en P_s ; n_s : número de subparcelas efectivamente evaluadas en campo en la parcela s , $n_s=1,2,3,4$; a_k : área de la subzona de inclusión del elemento U_k ; $n_s a_k$: área de la zona de inclusión del elemento U_k ; $\pi_k = n_s a_k / A$: probabilidad de inclusión del elemento U_k ; y_k : valor de un atributo de interés asociado a U_k (área basal, volumen, biomasa, carbono); $T_y = \sum_{k=1}^N y_k$: total poblacional de un atributo de interés; N : número de árboles en el área forestal de la región de interés; t : número de árboles en la subparcela.

Estimadores

Bajo el principio de que los conglomerados pueden ser de tamaño diferente, Velasco *et al.* (2005) propusieron estimadores de razón de medias para analizar datos del Inventario Nacional Forestal de México. En el denominador consideraron el atributo de interés (volumen) y en denominador el área del conglomerado como variable auxiliar. En este sentido el estimador se expresa como:

$$y_{ha} = \frac{\sum_{s=1}^m y_s}{\sum_{s=1}^m a_s} = \frac{\sum_{s=1}^m \sum_{j=1}^{n_s} \sum_{k=1}^t y_{sjk}}{\sum_{s=1}^m n_s a_{sj}} \tag{1}$$

En (1) y_s es la cantidad total del atributo en la parcela muestral s y a_s es el área (ha) de la parcela muestral s efectivamente evaluada en campo. La varianza estimada de (1), según Cochran (1993), es:

$$v(y_{ha}) = \frac{1}{\bar{a}^2} \frac{1}{m(m-1)} \sum_{s=1}^m (y_s - y_{ha} a_s)^2 \tag{2}$$

donde $\bar{a} = \sum_{s=1}^m a_s / m$. El estimador (1) funciona bien si: 1) la población completa a ser inventariada es teselada por todas las

(x_s, z_s) : coordinates of the cluster P_s ; (x_k, z_k) : geographic location of an element of the population, U_k ; a_{sj} : area of each subplot (400 m^2) in P_s ; n_s : number of subplots effectively evaluated in the field in plot s , $n_s=1,2,3,4$; a_k : area of the sub-zone of inclusion of the element U_k ; $n_s a_k$: area of the zone of inclusion of the element U_k ; $\pi_k = n_s a_k / A$: probability of inclusion of element U_k ; y_k : value of an attribute of interest associated with U_k (basal area, volume, biomass, carbon); $T_y = \sum_{k=1}^N y_k$: population total of an attribute of interest; N : number of trees in the forested area of the region of interest; t : number of trees in the subplot.

Estimators

Under the principle that clusters can be of different sizes, Velasco *et al.* (2005) proposed ratio of means estimators to analyze data of the National Forest Inventory of Mexico. In the numerator, they considered the attribute of interest (volume) and in the denominator the area of the cluster as the auxiliary variable. In this sense, the estimator is expressed as follows:

$$y_{ha} = \frac{\sum_{s=1}^m y_s}{\sum_{s=1}^m a_s} = \frac{\sum_{s=1}^m \sum_{j=1}^{n_s} \sum_{k=1}^t y_{sjk}}{\sum_{s=1}^m n_s a_{sj}} \tag{1}$$

In (1), y_s is the total quantity of the attribute in sampling plot s and a_s is the area (ha) of the sampling plot s effectively evaluated in the field. The variance estimated in (1), according to Cochran (1993) is:

$$v(y_{ha}) = \frac{1}{\bar{a}^2} \frac{1}{m(m-1)} \sum_{s=1}^m (y_s - y_{ha} a_s)^2 \tag{2}$$

where $\bar{a} = \sum_{s=1}^m a_s / m$. The estimator (1) functions well if 1) the entire population to be inventoried is tessellated by all the possible sample plots (non-overlapping clusters with no spaces between), and 2) the plots are selected with equal probabilities. Normally, INFyS does not satisfy either assumption in practice. Instead, the sample plots are selected as a sample point of an infinite universe of possible central points of plots, and the trees present where there are different sized plots are selected with a higher probability than those in smaller plots (Smelko and Saborowski, 1999).

The ratio estimator is a biased estimator; however, there may be certain circumstances in which it can be unbiased. Ratio

posibles parcelas muestrales (conglomerados no traslapados y sin espacios entre ellos), y 2) las parcelas son seleccionadas con probabilidades iguales. Ambos supuestos normalmente no son cumplidos en la práctica del INFyS, y en su lugar las parcelas muestrales se seleccionan como un punto muestral de un universo infinito de posibles puntos centrales de parcelas, y los árboles presentes donde hay parcelas de tamaños diferentes son seleccionados con probabilidad más alta que aquellos de parcelas más pequeñas (Smelko y Saborowski, 1999).

El estimador de razón es un estimador sesgado, pero hay circunstancias donde puede ser insesgado. El estimador de razón (1) es más efectivo cuando: 1) la relación entre y_s y a_s es una línea recta que pasa por el origen, y 2) la varianza de y_s alrededor de esta recta es proporcional a a_s . De manera práctica, el uso de (1) se justifica si: 1) el coeficiente de correlación entre y_s y a_s es mayor a 0.5, y 2) el tamaño de muestra es mayor a 30 (Scheaffer *et al.*, 1987; Cochran, 1993).

Scott *et al.* (2005), en el marco del programa de Inventario Forestal y Análisis (FIA) de EE.UU., propusieron un estimador para atributos forestales similar a una media de razones, pero para la estimación en cada parcela muestral s se considera un área ajustada:

$$y_{ha}(s) = \frac{y_s}{a_s} = \frac{\sum_{j=1}^{n_s} \sum_{k=1}^t y_{sjk}}{a_0 \bar{p}_0} \quad (3)$$

donde, a_0 es área total usada para observar el atributo de interés en una parcela (cuatro veces el área de la subparcela) y \bar{p}_0 es la proporción media de áreas de las parcelas observadas que caen dentro de la población de interés, se obtiene con

$$\bar{p}_0 = \sum_{s=1}^m \sum_{j=1}^{n_s} a_{sj} / a_0 m.$$

Los valores $y_{ha}(s)$ de (3) se usan para la estimación de la media por hectárea:

$$\bar{y}_{ha} = \frac{\sum_{s=1}^m y_{ha}(s)}{m} \quad (4)$$

La varianza de (4) se estima de la siguiente manera:

$$v(\bar{y}_{ha}) = \frac{\sum_{s=1}^m (y_{ha}(s) - \bar{y}_{ha})^2}{m(m-1)} \quad (5)$$

estimator (1) is more effective when 1) the relationship between y_s and a_s is a straight line that passes through the origin, and 2) the variance of y_s around this line is proportional to a_s . In practice, the use of (1) is justified if 1) the coefficient of correlation between y_s and a_s is above 0.5, and 2) the sample size is above 30 (Scheaffer *et al.*, 1987; Cochran, 1993).

Scott *et al.* (2005), within the framework of the Forest Inventory and Analysis (FIA) program of the United States, proposed an estimator for forest attributes similar to a mean of ratios, but for the estimation in each sample plot s , an adjusted area is considered:

$$y_{ha}(s) = \frac{y_s}{a_s} = \frac{\sum_{j=1}^{n_s} \sum_{k=1}^t y_{sjk}}{a_0 \bar{p}_0} \quad (3)$$

where a_0 is the total area used to observe the attribute of interest in a plot (four times the area of the subplot) and \bar{p}_0 is the mean proportion of plot areas that fall within the population of interest, obtained with $\bar{p}_0 = \sum_{s=1}^m \sum_{j=1}^{n_s} a_{sj} / a_0 m$.

The values $y_{ha}(s)$ of (3) are used to estimate the mean per hectare:

$$\bar{y}_{ha} = \frac{\sum_{s=1}^m y_{ha}(s)}{m} \quad (4)$$

The variance of (4) is estimated in the following manner:

$$v(\bar{y}_{ha}) = \frac{\sum_{s=1}^m (y_{ha}(s) - \bar{y}_{ha})^2}{m(m-1)} \quad (5)$$

The estimator of Scott *et al.* (2005) is a mean of ratios estimator that considers, in each sample plot s , the complete area of the cluster adjusted by the mean proportion of the areas of the plots observed that fall within the population of interest.

The probability of including a U_k element in each sample point allows deriving another estimator from the viewpoint of repeated sampling. Thus, T_y can estimate with an unbiased Horvitz-Thompson (HT) estimator, that is:

$$\hat{T}_{y\pi_s} = \sum_{U_k \in P_s} \frac{y_k}{\pi_k} = A \sum_{U_k \in P_s} \frac{y_k}{n_s a_k} = A \sum_{U_k \in P_s} \rho_k = A \rho_s \quad (6)$$

El estimador de Scott *et al.* (2005) es un estimador de media de razones que en cada parcela muestral s considera el área del conglomerado completo ajustada por la proporción media de áreas de las parcelas observadas que caen dentro de la población de interés.

La probabilidad de incluir un elemento U_k en cada punto muestral permite derivar otro estimador desde el punto de vista de muestreo replicado. Así, T_y se puede estimar mediante un estimador insesgado Horvitz-Thompson (HT), es decir:

$$\hat{T}_{y\pi s} = \sum_{U_k \in P_s} \frac{y_k}{\pi_k} = A \sum_{U_k \in P_s} \frac{y_k}{n_s a_k} = A \sum_{U_k \in P_s} \rho_k = A \rho_s \quad (6)$$

En la expresión (6) $\rho_k = y_k / n_s a_k$ es el valor de y_k prorrateado por unidad de área y ρ_s es la suma de todos los valores prorrateados para el conglomerado P_s . La instalación de múltiples conglomerados independientes P_s en A , seguido por la estimación HT de T_y con los datos de cada P_s (ecuación 6), es el muestreo replicado (Barabesi y Fattorini, 1998; Barabesi y Pisani, 2004). Según Schreuder *et al.* (1993) y Gregoire y Valentine (2008), el estimador de T_y basado en un muestreo replicado de $m P_s$ es:

$$\hat{T}_{y\pi,rep} = \frac{1}{m} \sum_{s=1}^m \hat{T}_{y\pi s} = \frac{1}{m} \sum_{s=1}^m A \rho_s = \frac{A}{m} \sum_{s=1}^m \rho_s = A \bar{\rho} \quad (7)$$

La varianza de $\hat{T}_{y\pi,rep}$ es $V(\hat{T}_{y\pi,rep}) = V(\hat{T}_{y\pi s}) / m$ (Barabesi y Pisani, 2004). Sin embargo, la varianza estimada de $\hat{T}_{y\pi s}$, derivada en términos del estimador HT, es algo tediosa para calcular, ya que requiere la probabilidad conjunta de seleccionar U_k y $U_{k'}$. Por esta razón, Särndal *et al.* (1992-pág. 424), Schreuder *et al.* (1993-pág. 116) y Gregoire y Valentine (2008-pág. 216) indican que $V(\hat{T}_{y\pi,rep})$ se puede estimar insesgradamente mediante:

$$v(\hat{T}_{y\pi,rep}) = \frac{1}{m(m-1)} \sum_{s=1}^m (\hat{T}_{y\pi s} - \hat{T}_{y\pi,rep})^2 \quad (8)$$

La cantidad total de un atributo de interés dividido por el área de la población resulta en la cantidad promedio por unidad de área, es decir, $\lambda_y = T_y / A$. Esta cantidad puede ser estimada insesgradamente con datos de la muestra mediante $\hat{\lambda}_{y\pi s} = \hat{T}_{y\pi s} / A$. Considerando esta expresión y (7), el estimador de λ_y a partir de muestreo replicado de m conglomerados es:

$$\hat{\lambda}_{y\pi,rep} = \frac{1}{m} \sum_{s=1}^m \hat{\lambda}_{y\pi s} = \frac{1}{A} \frac{1}{m} \sum_{s=1}^m \hat{T}_{y\pi s} = \frac{1}{A} \hat{T}_{y\pi,rep} = \bar{\rho} \quad (9)$$

In equation (6), $\rho_k = y_k / n_s a_k$ is the value of y_k prorated per unit of area and ρ_s is the sum of all the prorated values for the P_s cluster. Installation of multiple independent P_s clusters in A , followed by the HT estimation of T_y with data from each P_s (equation 6) is replicated sampling (Barabesi and Fattorini, 1998; Barabesi and Pisani, 2004). According to Schreuder *et al.* (1993) and Gregoire and Valentine (2008), the T_y estimator based on replicated sampling of $m P_s$ is:

$$\hat{T}_{y\pi,rep} = \frac{1}{m} \sum_{s=1}^m \hat{T}_{y\pi s} = \frac{1}{m} \sum_{s=1}^m A \rho_s = \frac{A}{m} \sum_{s=1}^m \rho_s = A \bar{\rho} \quad (7)$$

The variance of $\hat{T}_{y\pi,rep}$ is $V(\hat{T}_{y\pi,rep}) = V(\hat{T}_{y\pi s}) / m$ (Barabesi and Pisani, 2004). However, the estimated variance of $\hat{T}_{y\pi s}$ derived in terms of the HT estimator is somewhat tedious for computer since it requires the joint probability of selecting U_k and $U_{k'}$. For this reason, Särndal *et al.* (1992-p. 424), Schreuder *et al.* (1993-p. 116) and Gregoire and Valentine (2008-p. 216) point out that $V(\hat{T}_{y\pi,rep})$ can be estimated unbiasedly with:

$$v(\hat{T}_{y\pi,rep}) = \frac{1}{m(m-1)} \sum_{s=1}^m (\hat{T}_{y\pi s} - \hat{T}_{y\pi,rep})^2 \quad (8)$$

The total quantity of an attribute of interest divided by the area of the population results in the average quantity per unit of area; that is, $\lambda_y = T_y / A$. This quantity can be estimated unbiasedly with data from the sample with $\hat{\lambda}_{y\pi s} = \hat{T}_{y\pi s} / A$. Considering this expression and (7), the estimator of λ_y parting from the replicated sampling of m clusters is:

$$\hat{\lambda}_{y\pi,rep} = \frac{1}{m} \sum_{s=1}^m \hat{\lambda}_{y\pi s} = \frac{1}{A} \frac{1}{m} \sum_{s=1}^m \hat{T}_{y\pi s} = \frac{1}{A} \hat{T}_{y\pi,rep} = \bar{\rho} \quad (9)$$

where $\hat{T}_{y\pi s} = A \sum_{U_k \in P_s} (y_k / n_s a_k)$. The result (9), based on the prorated estimation per unit of area, $\rho_k = y_k / n_s a_k$, does not require knowledge of A nor of the explicit determination of inclusion probabilities. The variance of $\hat{\lambda}_{y\pi,rep}$ is $V(\hat{\lambda}_{y\pi,rep}) = V(\hat{T}_{y\pi,rep}) / A^2$, which, according to Gregoire and Valentine (2008-p. 220), can be estimated unbiasedly by:

$$v(\hat{\lambda}_{y\pi,rep}) = \frac{1}{m(m-1)} \sum_{s=1}^m (\hat{\lambda}_{y\pi s} - \hat{\lambda}_{y\pi,rep})^2 \quad (10)$$

Equation (9) does not require explicit determination of the inclusion probabilities of each tree, but it does assume that regardless of the number of satellite subplots that finally integrate

donde, $\hat{T}_{y\pi s} = A \sum_{U_k \in P_s} (y_k / n_s a_k)$. El resultado (9), basado en la estimación prorrateada por unidad de área, $\rho_k = y_k / n_s a_k$, no requiere conocer A ni la determinación explícita de las probabilidades de inclusión. La varianza de $\hat{\lambda}_{y\pi,rep}$ es $V(\hat{\lambda}_{y\pi,rep}) = V(\hat{T}_{y\pi,rep}) / A^2$, la cual, según Gregoire y Valentine (2008-pág. 220), puede ser estimada insesgadamente por:

$$v(\hat{\lambda}_{y\pi,rep}) = \frac{1}{m(m-1)} \sum_{s=1}^m (\hat{\lambda}_{y\pi s} - \hat{\lambda}_{y\pi,rep})^2 \quad (10)$$

La ecuación (9) no requiere la determinación explícita de las probabilidades de inclusión de cada árbol, pero sí supone que independiente del número de suparcelas satélite que al final integren el conglomerado en campo, las subzonas de inclusión estén completas (no truncadas). Así, (9) implica que para cada elemento, el atributo de interés sea dividido entre el área del conglomerado efectivamente evaluado en campo ($n_s a_k = n_s a_{sj}$), la cual es común para todos los árboles que pertenecen al mismo conglomerado.

Suponiendo que las subzonas de inclusión de los árboles estén completas, el estimador (9) puede considerarse como una media de razones (Smelko y Merganic, 2008). El método media de razones fue recomendado para muestreo sistemático con parcelas muestrales de tamaño diferente (Smelko y Saborowski, 1999). Por lo tanto, en cada parcela muestral s los datos muestrales y_s necesitan ser calculados a un área igual (1 ha) usando la siguiente formula (Smelko y Merganic, 2008):

$$y_{ha}(s) = \frac{y_s}{a_s} = \frac{\sum_{j=1}^{n_s} \sum_{k=1}^t y_{sjk}}{n_s a_{sj}} \quad (11)$$

donde y_s es la cantidad total del atributo en la parcela muestral s y a_s es el área (ha) de la parcela muestral s efectivamente evaluada en campo.

Estos valores por hectárea $y_{ha}(s)$ son usados para la estimación de la media por hectárea:

$$\bar{y}_{ha} = \frac{\sum_{s=1}^m y_{ha}(s)}{m} \quad (12)$$

No existen estimadores para la varianza de (12) cuando se utiliza muestreo sistemático, por lo tanto se estima de la siguiente manera:

the cluster in the field, the sub-zones of inclusion are complete (not truncated). Thus, (9) implies that for each element, the attribute of interest is divided by the area of the cluster effectively evaluated in the field ($n_s a_k = n_s a_{sj}$), which is common for all the trees belonging to the same cluster.

Assuming that the sub-zones of tree inclusion are complete, estimator (9) can be considered a mean of ratios (Smelko and Merganic, 2008). The method mean of ratios was recommended for systematic sampling with sample plots of different sizes (Smelko and Saborowski, 1999). Therefore, in each sample plot s , the sample data y_s need to be calculated to an equal area (1 ha) using the following formula (Smelko and Merganic, 2008):

$$y_{ha}(s) = \frac{y_s}{a_s} = \frac{\sum_{j=1}^{n_s} \sum_{k=1}^t y_{sjk}}{n_s a_{sj}} \quad (11)$$

where y_s is the total quantity of the attribute in the sample plot s and a_s is the area (ha) of the sample plot s effectively evaluated in the field.

These values per hectare $y_{ha}(s)$ are used to estimate the mean per hectare:

$$\bar{y}_{ha} = \frac{\sum_{s=1}^m y_{ha}(s)}{m} \quad (12)$$

There are no estimators for the variance of (12) when systematic sampling is used. It is therefore estimated in the following way:

$$v(\bar{y}_{ha}) = \frac{\sum_{s=1}^m (y_{ha}(s) - \bar{y}_{ha})^2}{m(m-1)} \quad (13)$$

this estimation was shown as a conservative variance in applications of systematic sampling (Smelko and Saborowski, 1999).

Estimation of the variance for the estimators (1), (4) and (12) considers random sampling even when the sampling design of the National Forest Inventory is systematic. However, in forest inventories, it is reasonable to assume that systematic sampling is approximately equivalent to simple random sampling (Zarnoch and Bechtold, 2000). Velasco *et al.* (2005) refer to the use of random sampling in forest inventories when the data are collected in a systematic grid. Cochran (1993) states that if the

$$v(\bar{y}_{ha}) = \frac{\sum_{s=1}^m (y_{ha}(s) - \bar{y}_{ha})^2}{m(m-1)} \quad (13)$$

la cual se mostró que es una varianza conservadora en aplicaciones de muestreo sistemático (Smelko y Saborowski, 1999).

La estimación de la varianza para los estimadores (1), (4) y (12) considera un muestreo aleatorio aun cuando el diseño de muestreo del inventario nacional forestal es sistemático. Sin embargo, en inventarios forestales es razonable suponer que el muestreo sistemático es aproximadamente equivalente al muestreo aleatorio simple (Zarnoch y Bechtold, 2000). Velasco *et al.* (2005) refieren el uso del muestreo aleatorio en inventarios forestales cuando los datos son recolectados en una malla sistemática. Cochran (1993) indica que si la población es aleatoria la fórmula para la varianza en muestreo sistemático es la misma que para una muestra aleatoria simple.

Otra forma para estimar la varianza de (1), (4) y (12) es con el método bootstrap no paramétrico, en el cual la varianza se calcula con base en un gran número de estimaciones correspondientes a muestras bootstrap obtenidas desde la muestra original (Efron, 1979; Särndal *et al.*, 1992; Pérez, 2000). En cualquier caso, suponiendo una distribución de $\hat{\theta}$ aproximadamente normal o Gaussiana, un intervalo de confianza para θ está dado de la siguiente manera: $\hat{\theta} \pm 2(v(\hat{\theta}))^{1/2}$.

Base de datos

En este estudio se usaron datos de selva mediana subperennifolia y subcaducifolia de Quintana Roo y Campeche, tomados en campo en 2010 en el marco del proyecto de Inventario Nacional Forestal y de Suelos. Para Quintana Roo se usaron $m=206$ conglomerados de selva mediana subperennifolia y $m=22$ de selva mediana subcaducifolia; para Campeche $m=127$ conglomerados de selva mediana subperennifolia y $m=58$ de selva mediana subcaducifolia. Las ecuaciones usadas fueron las de volumen fustal reportadas por SAG (1976) y SARH (1985), las cuales son aplicación a nivel estatal y por grupo botánico. La elección del volumen maderable como parámetro de interés se basó en que a nivel regional existen funciones de volumen en el Sureste mexicano. Las variables volumen y carbono están altamente correlacionadas, por lo cual se esperan resultados similares cuando la variable carbono sea de interés. Para el análisis de la información se elaboró un programa en el software Sistema de Análisis Estadístico Versión 9.2 (SAS, 2009). En este programa se generó una matriz (mxp); sus columnas correspondieron a la longitud, latitud, área basal, densidad arbórea, volumen maderable fustal y área

population is random, the formula for the variance in systematic sampling is the same as for simple random sampling.

Another way to estimate the variance of (1), (4) and (12) is with the non-parametric bootstrap method; with this, the variance is calculated on the basis of a large number of estimations corresponding to bootstrap samples obtained from the original sample (Efron, 1979; Särndal *et al.*, 1992; Pérez, 2000). In any case, assuming for $\hat{\theta}$ an approximately normal, or Gaussian, distribution, a confidence interval for θ is given as follows: $\hat{\theta} \pm 2(v(\hat{\theta}))^{1/2}$.

Database

In this study, data from medium height sub-evergreen and sub-deciduous tropical forests of Quintana Roo and Campeche taken in the field in 2010 were used within the framework of the project National Forest and Soils Inventory. For Quintana Roo, $m=206$ clusters of medium height sub-evergreen tropical forest and $m=22$ medium height sub-deciduous tropical forest were used. For Campeche, the data included $m=127$ clusters of medium height sub-evergreen tropical forest and $m=58$ clusters of medium height sub-deciduous tropical forest. The equations used were those of trunk volume reported by SAG (1976) and SARH (1985), which were applicable at the state level and by botanical group. Timber volume was selected as the parameter of interest because at the regional level there are functions of volume in the Mexican Southeast. The variables volume and carbon are highly correlated, so that similar results are expected when the variable carbon is of interest. For analysis of the information, a program was created with the software Statistical Analysis System Version 9.2 (SAS, 2009). In this program, a matrix (mxp) was constructed. Columns correspond to longitude, latitude, basal area, tree density, trunk timber volume and area of the cluster effectively sampled; rows identified the clusters.

The medium height sub-evergreen and sub-deciduous tropical forest of the Mexican southeast is of little importance from the quantitative perspective *per se*, but we decided to conduct this study in four of these tropical forest populations in Quintana Roo and Campeche to evaluate the statistical behavior of the estimators with different sample sizes. In these conditions, we hope the results will be generalizable for other forest populations of interest.

The timber volume per hectare was estimated with the following methods: 1) Mean of ratios (Smelko and Merganic, 2008), 2) ratio of means (Velasco *et al.*, 2005) and 3) mean of ratios with adjusted area (Scott *et al.*, 2005). The variances were obtained according to the conventional expressions reported for each estimator. Besides the point

efectivamente muestreada del conglomerado, y sus filas identificaron a los conglomerados.

La selva mediana subperennifolia y subcaducifolia del sureste mexicano tiene importancia limitada desde el punto de vista cuantitativo per se, pero se decidió realizar la presente investigación en cuatro poblaciones forestales tropicales diferentes —selva mediana subperennifolia y subcaducifolia de Quintana Roo y Campeche— para evaluar el comportamiento estadístico de los estimadores bajo distintos tamaños de muestra. En estas condiciones se espera que los resultados sean generalizables a otras poblaciones de interés forestal.

El volumen maderable fustal por hectárea se estimó con los siguientes métodos: 1) Media de razones (Smelko y Mer-ganic, 2008), 2) razón de medias (Velasco *et al.*, 2005) y, 3) media de razones con área ajustada (Scott *et al.*, 2005). Las varianzas se obtuvieron según las expresiones convencionales reportadas para cada estimador. Además de las estimaciones puntuales se obtuvieron las estimaciones por intervalo. Para el estimador de razón adicionalmente se estimó el sesgo median-

te $\hat{B}(y_{ha}) = \frac{1}{m\bar{a}^2} (y_{ha}s_a^2 - s_{y,a})$, donde $s_a^2 = \frac{\sum_{s=1}^m (a_s - \bar{a})^2}{m-1}$ y $s_{y,a} = \frac{\sum_{s=1}^m (y_s - \bar{y})(a_s - \bar{a})}{m-1}$ (Pérez, 2000). El sesgo relativo se obtuvo como $(\hat{B}(y_{ha}) / [v(y_{ha})]^{0.5}) \times 100$.

El traslape de los intervalos de confianza y similitud en el error de muestreo se utilizaron como criterios para comparar los estimadores estudiados.

RESULTADOS Y DISCUSIÓN

La correlación entre el volumen maderable y el área del conglomerado efectivamente muestreada en campo fue alrededor del 0.43. En las condiciones estudiadas las correlaciones fueron significativas al 5 %.

Los diagramas de dispersión indicaron que en todos los casos la varianza del volumen fue proporcional al área del conglomerado. En la selva mediana subperennifolia (Quintana Roo y Campeche) los conglomerados pequeños (0.04, 0.08 y 0.12 ha) estuvieron bien representados, pero en la selva mediana subcaducifolia sucedió lo contrario. Así, en Quintana Roo ($m=22$) no hubo conglomerados con área de 0.12 ha, y en Campeche ($m=58$) solo un conglomerado ocurrió con área de 0.04 ha y otro con área de 0.08 ha. Esta última situación puede deberse al tamaño de muestra relativamente bajo.

estimations, estimations by interval were also obtained. Additionally, for the ratio estimator, bias was estimated with

$\hat{B}(y_{ha}) = \frac{1}{m\bar{a}^2} (y_{ha}s_a^2 - s_{y,a})$, where $s_a^2 = \frac{\sum_{s=1}^m (a_s - \bar{a})^2}{m-1}$ and $s_{y,a} = \frac{\sum_{s=1}^m (y_s - \bar{y})(a_s - \bar{a})}{m-1}$ (Pérez, 2000). The relative bias was obtained as $(\hat{B}(y_{ha}) / [v(y_{ha})]^{0.5}) \times 100$.

The overlap of the confidence intervals and the similarity in the sampling error were used as criteria to compare the estimators studied.

RESULTS AND DISCUSSION

The correlation between timber volume and cluster area effectively sampled in the field was around 0.43. Under the conditions studied, the correlations were significant at a 5 %.

The dispersion diagrams indicated that, in all cases, the variance of volume was proportional to the area of the cluster. In the medium height sub-evergreen tropical forest (Quintana Roo and Campeche) the small clusters (0.04, 0.08 and 0.12 ha) were well represented. The opposite occurred in the medium sub-deciduous tropical forest. Thus in Quintana Roo ($m=22$) there were no clusters with an area of 0.12 ha, and in Campeche ($m=58$) there was only one with an area of 0.04 ha and another with an area of 0.08 ha. The latter situation may have been due to the relatively small sample size.

The regressions of volume over area of the cluster, $y = \beta_0 + \beta_1 a$ indicated that, for the medium height sub-evergreen tropical forest of Quintana Roo and Campeche, the intercept was significant at 5 % (value of p of the t test below 0.05); that is, $H_0: \beta_0 = 0$ was rejected. In the medium height sub-deciduous tropical forest of the two states, $H_0: \beta_0 = 0$ was not rejected. Given the dispersion of the observations, in no case did the coefficient of determination was higher than 0.5.

The above situation, puts the ratio of means estimator at a disadvantage. To be effective it requires a coefficient of correlation between the variable of interest and the auxiliary variable above 0.5, and the relationship must be a straight line that passes through the origin. However, it must be kept in

Las regresiones del volumen sobre el área del conglomerado, $y = \beta_0 + \beta_1 a$, indicaron que para la selva mediana subperennifolia de Quintana Roo y Campeche el intercepto fue significativo al 5 % (valor de p de la prueba de t menor a 0.05), es decir, se rechazó $H_0: \beta_0 = 0$. En la selva mediana subcaducifolia, de las dos entidades federativas, no se rechazó $H_0: \beta_0 = 0$. Dada la dispersión de las observaciones, en ningún caso el coeficiente de determinación superó al 0.5.

La situación anterior pone en desventaja al estimador de razón de medias, que para ser efectivo requiere un coeficiente de correlación mayor al 0.5 entre la variable de interés y la variable auxiliar, y la relación entre ambas debe ser una línea recta que pase por el origen. Sin embargo, debe tenerse en mente que en tres de las cuatro poblaciones estudiadas el número de observaciones rebasa en mucho el tamaño de muestra sugerido, el cual debe ser por lo menos de 30; así el uso del estimador de razón se justifica.

Los valores puntuales por hectárea calculados mediante los tres estimadores, en cada población, también fueron parecidos, y de manera consistente, el estimador media de razones proporciona valores ligeramente menores (conservadores) en comparación con los otros dos estimadores. Los sesgos relativos estimados del estimador de razón (sesgo respecto al error estándar) fueron de 0.18, 0.73, 0.43 y 0.22 %, para SMSUPQROO, SMSUCQROO, SMSUPCAMP y SMSUCCAMP, respectivamente. El sesgo relativo de 0.73 % correspondió a la población muestreada con sólo 22 conglomerados. Todos estos valores son menores al 10 %, por lo que el sesgo del estimador de razón es prácticamente insignificante (Cuadro 1).

En todos los casos el estimador de razón de medias, a pesar de que en general existe correlación débil entre la variable de interés y la auxiliar, es ligeramente más preciso (menor error de muestreo), le siguen el estimador media de razones y FIA. Esto puede explicarse porque en la selva mediana subperennifolia el tamaño de muestra fue mayor a 100, y en el caso de la selva mediana subcaducifolia, aunque el tamaño de muestra fue menor a 100, hay evidencia de una relación lineal al origen del volumen y el área del conglomerado. Al existir traslape de los intervalos de confianza al 95 %, puede considerarse que los tres estimadores no son estadísticamente diferentes. También, en cada población estudiada, los errores de

mind that in three of the four populations studied, the number of observations greatly surpasses the suggested sample size, which should be at least 30; thus, use of the ratio estimator is justified.

The point values per hectare calculated with the three estimators in each population were also similar. Consistently, the mean of ratios estimator provides slightly lower values (conservative) compared with the other two estimators. The relative biases estimated by the ratio estimator (bias relative to the standard error) were 0.18, 0.73, 0.43 and 0.22 %, for SMSUPQROO, SMSUCQROO, SMSUPCAMP and SMSUCCAMP, respectively. The relative bias of 0.73 % corresponded to the sampled population with only 22 clusters. All of these values are below 10 %, and so the bias of the ratio estimator is practically insignificant (Table 1).

In all cases, the ratio of means estimator, even though there is generally a weak correlation between the variable of interest and the auxiliary, it is slightly more precise (lower sampling error), followed by the mean of ratios estimator and FIA. This may explain why the sample size was larger than 100 in the medium height sub-evergreen tropical forest. In the case of the medium height sub-deciduous tropical, although the sample size was smaller than 100, there is evidence of a linear relationship to the origin of volume and cluster area. Because there is an overlap of the 95 % confidence intervals, it can be considered that the three estimators are not statistically different. In addition, in each of the studied populations the sampling errors of the evaluated estimators had a high degree of similarity.

In the two populations of each state, the FIA estimator was slightly less precise than the other two. Consistently, it is observed that it has a larger sampling error, even when the sample size is larger than 100. The reason for this may be that the denominator of the expression that permits obtaining the values per hectare considers a common adjusted area; the area of the cluster originally desirable (0.16 ha) is adjusted by a mean proportion of all the areas effectively sampled in the field.

The above results are congruent with other studies that compared mean of ratios and ratio of means estimators in the field of forestry. In this respect, Smelko and Saborowski (1999), with data

Cuadro 1. Estimaciones puntuales y por intervalo del volumen maderable ($\text{m}^3 \text{ha}^{-1}$) en selva mediana subperennifolia y subcaducifolia en Quintana Roo y Campeche, México, obtenidas mediante los estimadores: Media de razones (MR), Razón de medias (RM) e Inventario Forestal y Análisis (FIA).**Table 1. Point estimation and by interval of timber volume ($\text{m}^3 \text{ha}^{-1}$) in medium height sub-evergreen and sub-deciduous tropical forests of Quintana Roo and Campeche, Mexico, obtained with the estimators Mean of ratios (MR), Ratio of means (RM) and Forest Inventory and Analysis (FIA).**

Población de interés	Estimador	Estimación puntual	Límite inferior	Límite superior	Error de muestreo (%)
SMSUPQROO ($m=206$)	MR	105.90	96.54	115.25	8.83
	RM	109.23	99.95	118.52	8.50
	FIA	109.23	99.26	119.21	9.13
SMSUCQROO ($m=22$)	MR	112.35	77.30	147.40	31.20
	RM	118.75	83.46	154.05	29.72
	FIA	118.75	80.00	157.50	32.63
SMSUPCAMP ($m=127$)	MR	106.10	94.32	117.87	11.10
	RM	111.50	99.85	123.14	10.44
	FIA	111.50	98.19	124.81	11.94
SMSUCCAMP ($m=58$)	MR	98.32	86.10	110.55	12.43
	RM	99.92	87.68	112.17	12.25
	FIA	99.92	86.78	113.07	13.15

SMSUPQROO: Selva mediana subperennifolia de Quintana Roo, SMSUCQROO: Selva mediana subcaducifolia de Quintana Roo, SMSUPCAMP: Selva mediana subperennifolia de Campeche, SMSUCCAMP: Selva mediana subcaducifolia de Campeche ❖ SMSUPQROO: medium height sub-evergreen tropical forest of Quintana Roo, SMSUCQROO: Medium height sub-deciduous tropical forest of Quintana Roo, SMSUPCAMP: medium height sub-evergreen tropical forest of Campeche, SMSUCCAMP: Medium height sub-deciduous tropical forest of Campeche.

muestreo de los estimadores evaluados presentaron un grado alto de similaridad.

En las dos poblaciones estudiadas de cada estado, el estimador FIA tuvo una precisión un poco más baja que los otros dos; de manera consistente se observa que presenta mayor error de muestreo incluso donde el tamaño de muestra es superior a 100. Esto puede deberse a que el denominador de la expresión que permite obtener los valores por hectárea considera un área ajustada común; el área del conglomerado originalmente deseable (0.16 ha) es ajustada por una proporción media de todas las áreas efectivamente muestreadas en campo.

Los resultados anteriores son congruentes con otros estudios que compararon estimadores media de razones y razón de medias en el ámbito forestal. Al respecto, Smelko y Saborowski (1999) con datos de la región forestal del norte de Slovakia, 32 parcelas muestrales de diferente tamaño (desde 100 hasta 1000 m^2 en intervalos de 100) levantadas en 1991 y 1995, estudiaron los estimadores media de razones (método A) y razón de medias (método B) en la estimación del número de árboles y volumen

from the forested region of northern Slovakia, 32 sample plots of different sizes (from 100 to 1000 m^2 in ranges of 100) collected in 1991 and 1995, studied the mean of ratios (method A) and ratio of means estimators (method B) in the estimation of number of trees and timber volume per hectare. The correlations between number of trees and the area of the plots were low (0.323 and 0.278 for 1991 and 1995, respectively), whereas the correlations between volume and area of the plots were small (0.323 and 0.278 for 1991 and 1995, respectively). In the second case, no relationship of a straight line passing through the origin was observed. In both years, method A gave a lower volume per hectare than B (1991: A 423.3 and B 453.9 $\text{m}^3 \text{ha}^{-1}$; 1995: A 443.5 and B 470 $\text{m}^3 \text{ha}^{-1}$) and standard errors were also lower than those of B. The opposite was true for the number of trees per hectare. On the two occasions and for the two parameters, the differences in standard errors in both methods were not above 2.7 %.

Smelko and Merganic (2008) evaluated the ratio of means and the mean of ratios estimators with data

maderable por hectárea. Las correlaciones entre el número de árboles y el área de las parcelas fueron bajas (0.323 y 0.278 para 1991 y 1995, respectivamente), mientras que las correlaciones entre el volumen y el área de las parcelas fueron moderadas (0.82 ambas en ocasiones); en este segundo caso no se observó una relación de línea recta que pase por el origen. En los dos años el método A dio menor volumen por hectárea que el B (1991: A 423.3 y B 453.9 m³ ha⁻¹; 1995: A 443.5 y B 470 m³ ha⁻¹), y errores estándar más pequeños que el B. Lo contrario fue verdadero para el número de árboles por hectárea. Para las dos ocasiones y los dos parámetros las diferencias de los errores estándar en ambos métodos no superaron el 2.7 %.

Smelko y Merganic (2008) evaluaron el estimador de razón de medias y media de razones con datos del Inventario Nacional Forestal y Monitoreo de Slovakia. Para el volumen maderable comercial reportan los siguientes resultados por hectárea: $y_{ha,RM}=266.2 \text{ m}^3$, $(v(y_{ha,RM}))^{0.5} = 5.15 \text{ m}^3$, $y_{ha,MR}=263.9 \text{ m}^3$ y $(v(y_{ha,MR}))^{0.5} = 5.16 \text{ m}^3$. Esto conduce a diferencias en valores promedio por ha de 2.3 m³ y en error estándar de 0.01 m³. Entonces, bajo las condiciones estudiadas en Slovakia los dos métodos son prácticamente iguales en cuanto a precisión. No obstante, el estimador media de razones tiene como ventaja la obtención inmediata de estimaciones por hectárea, lo cual permite evaluar la variabilidad espacial del atributo de interés en el área de estudio. Además la fórmula de la varianza es relativamente más sencilla que el estimador de razón de medias.

En el estimador de razón los valores de las unidades de muestreo se tratan de manera global, esto es, se obtiene el total de la variable de interés y el total de la variable auxiliar, y al final se obtiene el cociente de ambos. En el estimador media de razones, el primer nivel de inferencia es a nivel de hectárea (parcela) y después a nivel poblacional. Un método estadístico potente para analizar datos de inventario obtenidos con muestreo panelizado es Mínimos Cuadrados Generalizados, el cual considera la covarianza entre parcelas remedidas. Así, es preferible mantener el nivel de inferencia a nivel de hectárea.

El tamaño de muestra tiene un efecto notable en el error de muestreo. Por ejemplo, para el volumen estimado por FIA, cuando $m=206$ el error de muestreo máximo es 9.15 %, cuando $m=127$ el error

of the National Forest Inventory and Monitoring of Slovakia. For commercial timber volume, they report the following results per hectare: $y_{ha,RM}=266.2 \text{ m}^3$, $(v(y_{ha,RM}))^{0.5} = 5.15 \text{ m}^3$, $y_{ha,MR}=263.9 \text{ m}^3$ and $(v(y_{ha,MR}))^{0.5} = 5.16 \text{ m}^3$. This leads to differences in average values per hectare of 2.3 m³ and a standard error of 0.01 m³. Thus, under the conditions studied in Slovakia, the two methods are practically equal in terms of precision. Nevertheless, the mean of ratios estimator has the advantage of obtaining immediate estimation per hectare, permitting evaluation of the spatial variability of the attribute of interest in the study area. Moreover, the formula of the variance is simpler than that of the ratio of means estimator.

With the ratio estimator, the values of the sampling units are treated globally; that is, the total of the variable of interest and the total of the auxiliary variable are obtained, and finally, the quotient of both is obtained. With the mean of ratios estimator, the first level of inference is at the hectare (plot) level, then at the population level. One powerful statistical method for analyzing inventory data obtained by panelized sampling is Generalized Least Squares, which considers the covariance between re-measured plots. Thus, it is preferable to maintain the inference level at the hectare level.

Sample size has a notable effect on the sample error. Thus, for example, for the volume estimated by FIA when $m=206$, the maximum sampling error is 9.15 %; when $m=127$, the maximum sampling error is 12.35 %; when $m=58$, the maximum sampling error is 13.24 %; and when $m=22$, the maximum sampling error shoots up to 32.85 %. With 22 observations, in the ratio and mean of ratio estimators there was also a sampling error of around 30 %. This situation is not a reason for concern considering that INFyS is designed for the study of forest populations of great magnitude. Low sample sizes of these populations will require greater sampling intensity to achieve reliable estimations.

When considering that $\bar{a} = \sum_{s=1}^m a_s / m \Rightarrow \bar{a}m = \sum_{s=1}^m a_s$,

it is easy to verify that FIA and RM are equal, even when the clusters are incomplete. That is:

máximo de muestreo es 12.35 %, cuando $m=58$ el error máximo de muestreo es 13.24 % y cuando $m=22$ el error máximo de muestreo asciende hasta 32.85 %. Con 22 observaciones, en los estimadores de razón y media de razones también se obtuvo un error de muestreo alrededor del 30 %. Esta situación no deber ser preocupante si se considera que el IN-FyS está diseñado para estudiar poblaciones forestales de gran magnitud. Poblaciones con tamaños de muestreo bajos requerirán una mayor intensidad de muestreo para lograr estimaciones confiables.

Al considerar que $\bar{a} = \sum_{s=1}^m a_s / m \Rightarrow \bar{a}m = \sum_{s=1}^m a_s$ es fácil verificar que los estimadores FIA y RM son iguales, incluso si los conglomerados son incompletos, esto es:

$$\begin{aligned}
 FIA = \bar{y}_{ha} &= \frac{\sum_{s=1}^m y_{ha}(s)}{m} = \frac{\sum_{s=1}^m \frac{y_s}{\sum_{s=1}^m a_s}}{m} = \frac{\sum_{s=1}^m \frac{y_s}{\bar{a}m}}{m} \\
 &= \frac{\sum_{s=1}^m \frac{y_s}{\bar{a}}}{m} = \frac{\sum_{s=1}^m y_s}{\bar{a}m} = \frac{\sum_{s=1}^m y_s}{\sum_{s=1}^m a_s} = RM
 \end{aligned}$$

Esta situación hace suponer que en otras poblaciones forestales diferentes a las selvas, los estimadores MR, RM y FIA tendrán un comportamiento como el obtenido en este estudio.

CONCLUSIONES

En cada situación evaluada, el traslape de los intervalos de confianza al 95 % permite concluir que los estimadores evaluados son equivalentes; es poco más preciso el estimador de razón.

El tamaño de muestra tiene un efecto fuerte en la precisión de los estimadores. Así, en la población donde se estudiaron menos de 30 conglomerados, los errores de muestreo de los tres estimadores analizados ascienden hasta 30 % aproximadamente en la estimación del volumen maderable.

Todos los sesgos relativos del estimador de razón fueron menores al 10 %, por lo cual se considera prácticamente insignificante.

$$\begin{aligned}
 FIA = \bar{y}_{ha} &= \frac{\sum_{s=1}^m y_{ha}(s)}{m} = \frac{\sum_{s=1}^m \frac{y_s}{\sum_{s=1}^m a_s}}{m} = \frac{\sum_{s=1}^m \frac{y_s}{\bar{a}m}}{m} \\
 &= \frac{\sum_{s=1}^m \frac{y_s}{\bar{a}}}{m} = \frac{\sum_{s=1}^m y_s}{\bar{a}m} = \frac{\sum_{s=1}^m y_s}{\sum_{s=1}^m a_s} = RM
 \end{aligned}$$

This situation leads us to assume that in other forest populations different from tropical rainforests, the MR, RM and FIA estimators would behave as they did in this study.

CONCLUSIONS

In each of the situations tested, the overlap of the confidence intervals at 95 % leads us to conclude that the evaluated estimators are equivalent, the ratio estimator being slightly more precise.

Sample size has a strong effect on the precision of the estimators. Thus, in the population where fewer than 30 clusters were studied, the sampling errors of the three estimators analyzed ascend to approximately 30 % in the estimation of timber volume.

All of the relative biases of the ratio estimator were less than 10 %, and so considered practically insignificant.

—End of the English version—



LITERATURA CITADA

Barabesi, L., and L. Fattorini 1998. The use of replicated plot, line and point sampling for estimating species abundance and ecological diversity. *Environ. Ecol. Stat.* 5: 353-370.

Barabesi, L., and C. Pisani 2004. Steady-state ranked set sampling for replicated environmental designs. *Environmetrics* 15: 45-46.

Bechtold, W. A., and C. T. Scott. 2005. The forest inventory and analysis plot design. *In*: Bechtold W. A., and P. L. Patterson (eds). *The enhanced Forest Inventory and Analysis program-national sampling design and estimation procedures*. General Technical Report SRS-80. Forest Service, United States Department of Agriculture. Southern Research Station, Asheville, North Carolina. USA. pp: 27-42.

- Cochran, W. G. 1993. *Técnicas de Muestreo*. Compañía Editorial Continental. México. 513 p.
- CONAFOR. 2012. *Manual y procedimientos para el muestreo de campo: Re-muestreo 2012*. Secretaría de Medio Ambiente y Recursos Naturales-Comisión Nacional Forestal (CONAFOR). México. 136 p.
- Eaton, J. M., and D. Lawrence. 2009. Loss of carbon sequestration potential after several decades of shifting cultivation in the Southern Yucatán. *For. Ecol. Manage.* 258: 949-958.
- Efron, B. 1979. Bootstrap methods: another look at the Jackknife. *Annals Stat.* 7: 1-26.
- Gregoire, T. G., and H. T. Valentine. 2008. *Sampling Strategies for Natural Resources and the Environment*. Chapman & Hall/CRC. New York, USA. 474 p.
- McRoberts, R. E., W. A. Bechtold, P. L. Patterson, C. T. Scott, and G. A. Reams. 2005. The enhanced Forest Inventory and Analysis program of the USDA Forest Service: Historical perspective and announcement of statistical documentation. *J. For.* 103: 304-308.
- Pérez L., C. 2000. *Técnicas de Muestreo Estadístico. Teoría, Práctica y Aplicaciones Informáticas*. Alfaomega Grupo Editor. México. 603 p.
- Särndal, C. E., B. Swensson, and J. Wretman. 1992. *Model Assisted Survey Sampling*. Springer-Verlag. New York, USA. 694 p.
- SAS Institute Inc. 2009. *SAS/STAT 9.2. User's Guide, Second Edition*. SAS Institute Inc. North Carolina. United States America. 7886 p.
- Scheaffer, R. L., W. Mendenhall, y L. Ott. 1987. *Elementos de Muestreo*. Rendón S., G., y J. R. Gómez A. (trad). Grupo Editorial Iberoamérica. México. 321 p.
- SAG. 1976. *Inventario forestal del estado de Quintana Roo*. Publicación Número 41. Secretaría de Agricultura y Ganadería (SAG). México. 44 p.
- SARH. 1985. *Inventario forestal del estado de Campeche*. Publicación Especial Número 56. Secretaría de Agricultura y Recursos Hidráulicos (SARH). México. 119 p.
- Smelko, S., and J. Saborowski. 1999. Evaluation of variable size sampling plots for monitoring of forest condition. *J. For. Sci.* 45: 341-347.
- Smelko, S., and J. Merganic. 2008. Some methodological aspects of the National Forest Inventory and Monitoring in Slovakia. *J. For. Sci.* 54: 476-483.
- Scott, C.T., W. A. Bechtold, G. A. Reams, W. D. Smith, J. A. Westfall, M. H. Hansen, and G. G. Moisen. 2005. Sample-based estimators used by the Forest Inventory and Analysis national information management system. *In*: Bechtold, W. A., and P. L. Patterson (eds). *The enhanced Forest Inventory and Analysis program-national sampling design and estimation procedures*. General Technical Report SRS-80. Forest Service. United States Department of Agriculture. Southern Research Station, Asheville, North Carolina. USA. pp: 43-67.
- Schreuder, H.T., T. G. Gregoire, and G. B. Wood. 1993. *Sampling Methods for Multiresource Forest Inventory*. John Wiley & Sons, Inc. New York, USA. 446 p.
- Valentine, H. T., M. J. Ducey, J. H. Gove, A. Lanz, and D. L. R. Affleck. 2006. Corrections for cluster-plot slop. *For. Sci.* 52: 55-66.
- Velasco B., E., H. Ramírez M., F. Moreno S., y A. de la Rosa V. 2005. Estimadores de razón para el inventario nacional forestal de México. *Rev. Ciencia For. Méx.* 28: 23-43.
- Zarnoch, S. J., and W. A. Bechtold. 2000. Estimating mapped-plot forest attributes with ratios of means. *Can. J. For. Res.* 30: 688-697.