BOSQUE 26(1): 65-75, 2005 DOI: 10.4067/s0717-92002005000100007

Determinación del tamaño de la muestra en el muestreo RBS con selección sin reposición en la primera etapa

Determining sample size in randomized branch sampling with selection without replacement during the first stage.

JORGE CANCINO

Facultad de Ciencias Forestales, Universidad de Concepción, Chile.
E-mail: jcancino@udec.cl

SUMMARY

Randomized Branch Sampling (RBS) is useful for the estimation of foliar or root biomass, amount of fruit, and other tree parameters. RBS uses the natural branching pattern within the crown of the tree to configure the sample. Its use requires the definition of nodes (those points where a branch or its part divides in two), segments (those parts of a branch between two consecutive nodes), and paths (those series of successive segments between the basal and a terminal segment, i.e., a segment without a node at its final end). Currently, three variants of the RBS exist. One corresponds to the traditional RBS, which applies selection with replacement in all the nodes; the other two variants apply selection without replacement in the first or the second node defined in the tree. The formula for the estimation of variance of the estimator in the traditional RBS is simple; thus, it is easy to obtain the formula to determine the sample size (number of paths) required to obtain a specific sampling error. Conversely, in the variants of the RBS with selection without replacement, the formulas for variance estimation of the estimator are complex and, until now, no formula exists for the determination of sample size. In this article, a graphical approach for the determination of sample size (number of primary segments and number of paths) is developed for RBS sampling with selection without replacement taking place in the first stage.

Keywords: sample size, selection without replacement, pps-selection.

RESUMEN

El muestreo aleatorizado de ramas (Randomized Branch Sampling: RBS) es útil para la estimación de parámetros tales como biomasa foliar, biomasa de raíces o cantidad de frutos en árboles. El RBS utiliza la ramificación natural dentro de la copa del árbol para configurar la muestra. Su uso requiere definir *nodos* (un punto donde una rama o parte de una rama se divide en dos o más ramas), *segmentos* (una parte de una rama entre dos nodos consecutivos), y *paths* (una serie de segmentos sucesivos entre el segmento basal y un segmento terminal, esto es un segmento sin nodo en su extremo final). En la actualidad existen tres variantes del RBS. La primera, que corresponde al RBS tradicional, aplica selección con reposición en todos los nodos; las otras dos variantes del método aplican selección sin reposición en el primer o segundo nodo definidos en el árbol.
La fórmula para la estimación de la varianza del estimador del RBS tradicional es bastante simple; así, es fácil obtener a partir de ésta la fórmula para determinar el tamaño de muestra (número de paths) requerido para lograr un error específico. Por el contrario, en las variantes del RBS que contemplan selección sin reemplazo, las fórmulas de estimación de la varianza del estimador son complejas, no existiendo hasta el momento una fórmula para la determinación del tamaño de muestra. En este artículo se desarrolla una aproximación gráfica para la determinación del tamaño de muestra (número de segmentos primarios y número de paths) en el muestreo RBS con selección sin reposición en la primera etapa.

Palabras clave: tamaño de muestra, muestreo sin reposición, selección con probabilidad proporcional al tamaño.

INTRODUCCION

El muestreo aleatorizado de ramas (Randomized Branch Sampling: RBS) fue desarrollado para la estimación de parámetros tales como biomasa foliar, biomasa de raíces o cantidad de frutos en árboles. El RBS utiliza la ramificación natural dentro de la copa para configurar la muestra. Su uso requiere definir *nodos* (un punto donde una rama o parte de una rama se divide en dos o más ramas), *segmentos* (una parte de una rama entre dos nodos consecutivos; figura 1a), y *paths* (una serie de segmentos sucesivos entre el segmento basal y un segmento terminal, esto es un segmento sin nodo en su extremo final).

La estimación de la variable de interés para el árbol completo se realiza a partir de los valores de la variable colectados a lo largo de uno o más paths. En la selección de los segmentos de un path puede usarse una variable auxiliar. La selección de un path se inicia en el primer nodo definido con la selección de uno de los segmentos que de allí emanan, continúa a lo largo del segmento seleccionado y repite la selección, si es que al final del segmento existe otro nodo. El path finaliza al seleccionar un segmento terminal (figura 1a).

El RBS es un método flexible. Existe libertad para definir tanto el primer nodo de un path y su último segmento, como también la variable auxiliar. El punto de inicio de cada path determina la parte del árbol para la cual es válida la estimación que el path provee. La elección de la variable auxiliar debe estar guiada por el objetivo de la estimación (1). Ejemplos de variable auxiliar son el área de sección transversal en la base de la rama (2, 3), la biomasa de hojas estimada visualmente (4), y el producto entre el diámetro al cuadrado y la longitud del segmento (1, 5).

La selección de los segmentos se realiza con probabilidad proporcional al tamaño de la variable auxiliar. Este aspecto está estrechamente relacionado con la precisión del estimador. Aunque en principio cualquier característica del segmento puede utilizarse como variable auxiliar, es recomendable seleccionar una variable auxiliar estrechamente relacionada con la variable objetivo, con el fin de obtener la mayor precisión posible (1, 2, 6, 7). Estratificar la copa del árbol (3, 4) y eliminar los segmentos gruesos (7) ayudan a aumentar la precisión de las estimaciones.

En la actualidad existen tres variantes del RBS. La primera, que corresponde al RBS tradicional desarrollado por Jessen (2), aplica selección con reposición (swr: selection with replacement) en todos los nodos (8, 9, 10, 11); las otras dos variantes del método, desarrolladas por Saborowski y Gaffrey (12), aplican selección sin reposición (swor: selection without replacement) en el primer o segundo nodo.

El uso de swr que el RBS clásico hace en cada nodo puede causar una pérdida de eficiencia. Por esa razón Saborowski y Gaffrey (12) sugirieron el uso de swor en el primer o segundo nodo. Ello se basa en el hecho bien conocido que con muestreo aleatorio simple swor es más eficiente que swr (13). Los autores seleccionaron el método de Sampford (14) para la selección con probabilidad variable y sin reposición y lo incorporaron en el estimador multietápico de Saborowski (15). El estimador combina swor y selección con probabilidad variable con el estimador Horvitz-Thompson (16).

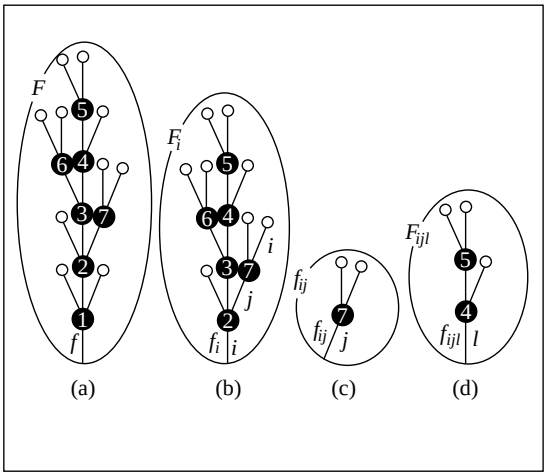

Figura 1: (a) Esquema de un árbol con 7 nodos y 17 segmentos. Los nodos 1 a 5 forman el fuste. (b), (c) y (d) representan 3 niveles de compartimentos de la copa, los que se inician en un segmento (*i*) primario, (*ij*) secundario y (*ijl*) terciario, con la correspondiente variable objetivo en los segmentos (f_i, f_{ij}, f_{ijl}), y los valores acumulados (F_i, F_{ij}, F_{ijl}).

(a) Scheme for a tree with seven nodes and 17 segments. Nodes 1–5 form the stem. (b), (c), and (d) represent three levels of compartments in the crown, which begin in a primary (*i*), secondary (*ij*), and tertiary (*ijl*) segment, with the corresponding objective variables in the segments (f_i, f_{ij}, f_{ijl}), and accumulated values (F_i, F_{ij}, F_{ijl}).

Descripción de los métodos RBS: Todas las variantes del RBS usan probabilidades de selección proporcional a una variable auxiliar, la que puede ser medida o estimada en los segmentos de un nodo. Así, la probabilidad de selección (*condicional*) del i^{mo} segmento en un nodo con N segmentos es dado por $q=x_i / \sum_{i=1}^{N} xi$, donde x_i es la variable auxiliar del i^{mo} segmento.

RBS clásico: En el RBS clásico, cada path seleccionado genera un estimador de la variable objetivo, la que se obtiene en base a los valores de la variable en cada segmento del path y la probabilidad (*incondicional*) del segmento respectivo. Si se mide p.e. un valor de la variable de interés f_r en el r^{mo} segmento del path, f_r / Q_r es la contribución de ese segmento al estimador de la variable de interés para la parte del árbol sobre el punto de inicio del path, con $\prod_{s=1}^{r} q_s$, donde q_s es la probabilidad de selección condicional del s^{mo} segmento del path. Así, el estimador de la variable objetivo total en el árbol, obtenido a partir de un path con R_0 segmentos que se inicia en el segmento basal del árbol, es:

$$\hat{F}_0 = \sum_{s=1}^{R_0} \frac{f_s}{Q_s} \qquad [1]$$

Si se seleccionan n paths mediante el RBS clásico, se obtiene el estimador insesgado $\bar{\hat{F}}$, $\bar{\hat{F}} = \frac{1}{n}\sum_{i=1}^{n} \hat{F}_{0i}$, con $\hat{F}_{0i} = \sum_{s=1}^{R_{0i}} \frac{f_s}{Q_s}$, y $Q_s = \prod_{u=1}^{s} q_u$, en donde el subíndice "0" se ha incorporado para resaltar que el estimador se obtiene desde un path que se inicia en el segmento basal del árbol, el único segmento perteneciente a un hipotético nodo cero. La varianza y el estimador insesgado de la varianza, respectivamente, son:

$$Var\bar{\hat{F}} = \frac{1}{n}\sum_{i=1}^{N_{Paths}} Q_{Ri}\left(\hat{F}_{0i} - F\right)^2, \text{ con } Q_{R_i} = \prod_{s=1}^{R_{0i}} q_s, \qquad [2]$$

y

$$V = \frac{1}{n(n-1)}\sum_{i=1}^{n}\left(\hat{F}_{0i} - \bar{\hat{F}}\right)^2, \qquad [3]$$

donde R_{0i} es el número de segmentos del i^{mo} path y N_{Paths} es el número total de paths posibles en el árbol.

En la práctica, el RBS tradicional no presenta ninguna limitación. Puede ser aplicado con cualquier tamaño de muestra en árboles enteros o en partes del árbol, considerando o no los segmentos del fuste principal como unidades elegibles. Al menos dos paths deben elegirse para estimar la varianza del estimador.

El RBS clásico es obviamente un muestreo aleatorio multietápico (12). A excepción del segmento basal, cada segmento del path puede ser asignado a cierta etapa. La primera selección corresponde a la primera etapa, y los segmentos que ramifican desde el primer nodo corresponden a unidades primarias; la segunda selección corresponde a la segunda etapa, y los segmentos pertinentes a las unidades secundarias, etc. Así, un nodo es un punto de transición desde un segmento a los segmentos de la etapa siguiente y el path es una secuencia de segmentos de diferentes etapas (figura 1).

En el muestreo aleatorio multietápico usual, la población está constituida de N unidades primarias, M_i unidades secundarias en la i^{ma} unidad primaria, K_{ij} unidades terciarias en la unidad secundaria j de la unidad primaria i, etc. Las probabilidades condicionales de selección son q_i, q_{ij}, q_{ijl}, etc. Los tamaños muestrales en las diferentes etapas son n, m_i, k_{ij}, etc. Concordantemente, los valores de la variable de interés de los segmentos sucesivos de un path f, f_1, f_2, f_3,..., pueden ser representados por f, f_i, f_{ij}, f_{ijl},..., y por F, F_i, F_{ij}, F_{ijl},..., los valores acumulados de la variable (figura 1b-d), donde F es el total de la variable de interés en el árbol, F_i es el total acumulado en la unidad primaria i, F_{ij} el de la unidad secundaria j en la unidad primaria i, F_{ijl} el de la unidad terciaria l de la unidad secundaria j en la unidad primaria i, etc. Así:

$$F = f + \sum_{i=1}^{N} F_i, \quad F_i = f_i + \sum_{j=1}^{M_i} F_{ij}, \quad F_{ij} = f_{ij} + \sum_{l=1}^{K_{ji}} F_{ijl}, \text{ etc.}$$

El RBS clásico selecciona n unidades con reposición (wr: with replacement) en la primera etapa, y $m_i = k_{ij} = 1$, es decir, por cada segmento primario de rama (unidad primaria) en todas las etapas siguientes se selecciona sólo un segmento de rama. Otra diferencia con los métodos multietápicos de selección aleatoria tradicionales yace en la composición de la variable de interés, como ya se vio antes. Aquí no sólo las unidades en la

última etapa sino también las unidades de todas las etapas anteriores pueden contribuir a la variable (ver relación [1]).

Un estimador insesgado de F es, por lo tanto:

$$\bar{\hat{F}} = f + \frac{1}{n}\sum_{i=1}^{n}\frac{\hat{F}_i}{q_i}, \text{ con, } \hat{F}_i = f_i + \frac{Y_{ij}}{q_{ij}} \qquad [4]$$

con varianza

$$Var\bar{\hat{F}} = \frac{1}{n}\sum_{i=1}^{N} q_i\left(\frac{F_i}{q_i} - \sum_{i=1}^{N}F_i\right)^2 + \frac{1}{n}\sum_{i=1}^{N}\frac{Var_2\hat{F}_i}{q_i}, \qquad [2']$$

donde $Var_2\hat{F}_i$ es la varianza condicional de $\hat{F}_i$ (estimador insesgado de F_i), dada la selección muestral en la primera etapa.

El estimador de la varianza [2'] es:

$$V = \frac{1}{n(n-1)}\sum_{i=1}^{n}\left(\frac{\hat{F}_i}{q_i} - \frac{1}{n}\sum_{i=1}^{n}\frac{\hat{F}_i}{q_i}\right)^2. \qquad [3']$$

Aunque la estructura de la fórmula para la determinación de la varianza del estimador de la característica de interés en el árbol [2'] es más compleja con la nueva nomenclatura (compare con [2]), el estimador de la varianza [3'], sin embargo, es equivalente a [3]. Note que $\hat{F}_{0i} = f + \frac{\hat{F}_i}{q_i}$ y, $\bar{\hat{F}} = f + \frac{1}{n}\sum_{i=1}^{n}\frac{\hat{F}_i}{q_i}$, de modo que $\hat{F}_{0i} - \bar{\hat{F}} = \frac{\hat{F}_i}{q_i} - \frac{1}{n}\sum_{i=1}^{n}\frac{\hat{F}_i}{q_i}$, lo que evidencia que el RBS es un tipo particular de muestreo aleatorio multietápico.

Aunque la nueva nomenclatura es más compleja y requiere mantener estimadores por unidad primaria, la introducción del RBS clásico en la nomenclatura de muestreo multietápico es necesaria para la incorporación de la selección sin reposición de unidades muestrales en alguna etapa de selección. En ese sentido es también útil notar que $\hat{F}_i$, un estimador del total de la variable de interés en la i^{ma} unidad primaria, puede obtenerse a partir del RBS clásico con un path que se inicie en la i^{ma} unidad primaria, esto es mediante $\hat{F}_{1i} = \sum_{s=1}^{R_{1i}}\frac{f_s}{Q_s}$, en el cual el subíndice "1" se ha incorporado para resaltar que el estimador se obtiene desde un path que se inicia en una unidad primaria, con $q_i = 1$. Así, [2'] se transforma en:

$$Var\bar{\hat{F}} = \frac{1}{n}\sum_{i=1}^{N} q_i\left(\frac{F_i}{q_i} - \sum_{i=1}^{N}F_i\right)^2 + \qquad [2'']$$

$$\frac{1}{n}\sum_{i=1}^{N}\frac{1}{q_i}\sum_{j=1}^{M_{Path.i}} Q_{R_{1ij}}\left(\hat{F}_{1ij} - F_i\right)^2$$

en la que $M_{Path.i}$ es el número total de paths que empiezan en la i^{ma} unidad primaria.

RBS con swor en la primera etapa (swor_swr): Una variante del método se obtiene usando swor en la primera etapa (12). Con este método se seleccionan n unidades sin reposición en la etapa 1 (wor: without replacement), $m_i \geq 1$ unidades wr en la etapa 2 en la unidad primaria i, y $k_{ij} = \ldots = 1$ unidades wr en todas las etapas siguientes. Así, reemplazando $\hat{F}_i$ por $\bar{\hat{F}}_{1i}$ [6] se obtiene el estimador formalmente idéntico a [4], cuya varianza y estimador de la varianza, ahora, respectivamente, son:

$$Var\bar{\hat{F}} = \sum_{i=1}^{N}\sum_{i'>i}^{N}(\pi_i\pi_{i'} - \pi_{ii'})\left(\frac{F_i}{\pi_i} - \frac{F_{i'}}{\pi_{i'}}\right)^2 + \qquad [5]$$

$$\sum_{i=1}^{N}\frac{1}{\pi_i}\frac{1}{m_i}\sum_{j=1}^{M_{Path.i}} Q_{R_{1ij}}\left(\hat{F}_{1ij} - F_i\right)^2$$

y

$$V = \sum_{i=1}^{n}\sum_{i'>i}^{n}\frac{\pi_i\pi_{i'} - \pi_{ii'}}{\pi_{ii'}}\left(\frac{\bar{\hat{F}}_{1i}}{\pi_i} - \frac{\bar{\hat{F}}_{1i'}}{\pi_{i'}}\right)^2 + \qquad [6]$$

$$\sum_{i=1}^{n}\frac{1}{\pi_i}\frac{1}{m_i(m_i-1)}\sum_{j=1}^{m_i}\left(\hat{F}_{1ij} - \bar{\hat{F}}_{1i}\right)^2$$

con $$\bar{\hat{F}}_{1i} = \frac{1}{m_i}\sum_{j=1}^{m_i}\hat{F}_{1ij},$$

en donde es evidente que el método requiere seleccionar al menos dos unidades primarias ($n \geq 2$) y al menos dos paths ($m_i \geq 2$) por unidad primaria para estimar la varianza del estimador.

La selección de las unidades, el cálculo de la probabilidad que la i^{ma} unidad sea incluida en la muestra (π_i), y la probabilidad que tanto las unidades i y i' estén en la muestra ($\pi_{ii'}$) se realizan con el método de Sampford (14). Con este método:

$$\pi_i = n \cdot q_i, \qquad [7]$$

y

$$\pi_{ii'} = K_n\lambda_i\lambda_{i'}\sum_{t=2}^{n}\frac{[t - n(q_i + q_{i'})]L_{n-t}(\bar{i}\bar{i}')}{n^{t-2}},$$

con $K_n = \left(\sum_{t=1}^{n} \frac{t L_{n-t}}{n^t} \right)^{-1}$ y $\lambda_i = \frac{q_i}{1 - n q_i}$. [8]

Sampford define como $S(m)$ a un conjunto de $m \leq N$ unidades i_1, i_2,..., i_m y define L_m como L_0=1, $L_m = \sum_{s(m)} \lambda_{i1} \lambda_{i2} \cdots \lambda_{im}$ $(1 \leq m \leq N)$, donde la suma incluye a todos los subconjuntos posibles de m unidades de la población; $L_m(\bar{i}\bar{i}')$ se define de manera similar, pero se obtiene de un población en la que se han eliminado las unidades i y i'.

DETERMINACION DEL TAMAÑO DE LA MUESTRA

Tamaño de la muestra en el RBS tradicional: En este caso, el tamaño de la muestra (número de paths a medir) se determina mediante:

$$n = \frac{1}{Var_e \bar{\hat{F}}} \sum_{i=1}^{N_{Paths}} Q_{Ri} \left(\hat{F}_{01} - F \right)^2, \qquad [9]$$

relación en la que $\overline{Var_e}^{\wedge}$ es la precisión esperada del estimador.

Tamaño de la muestra para la variante RBS con swor en la primera etapa: En esta variante del RBS, el máximo tamaño de muestra posible lo define la máxima probabilidad de selección en la primera etapa. Note que al ser $\pi_i = n \cdot q_i$ (relación [7]), y que, al tratarse de probabilidades, debe cumplirse $n \cdot q_i < 1$. Eso significa que el máximo tamaño de muestra lo define la máxima probabilidad de selección, esto es $n < 1/\max(q_i)$. Esto puede limitar severamente el tamaño de muestra, especialmente cuando el fuste principal se considera como un segmento más. El problema se reduce "eliminando" segmentos del fuste principal. Con ello se aumenta notoriamente el tamaño máximo de muestra posible para swor y se logra a su vez una reducción fuerte de la varianza del estimador (7). Todo segmento eliminado no participa en la elección de unidades; sin embargo, si en ellos existe algo de la variable de interés, ella debe medirse y al final agregarse al estimador manteniendo así la característica de insesgamiento. Al eliminar un segmento, el nodo localizado en su extremo se disuelve y todos sus segmentos son incorporados al nodo precedente; con ello aumenta el número de unidades en ese nodo y disminuye $\max(q_i)$, aumentando de ese modo el máximo tamaño de muestra posible.

La determinación analítica del tamaño muestral óptimo para el RBS con swor es complicada debido a lo complejo de la fórmula para calcular las probabilidades $\pi_{ii'}$ (relación [8]). Además, puede ocurrir que, de poder determinarse un tamaño n óptimo en forma analítica, éste puede exceder el máximo tamaño de muestra posible en la práctica, no justificando el esfuerzo destinado a ello. Así, es más práctico concentrarse en el análisis simultáneo de la precisión y costos de diferentes combinaciones de tamaños muestrales (n, m) en el rango de tamaños n posibles. En ello se basa el método que se propone a continuación para la determinación del tamaño de muestra.

El costo (tiempo) total C para recopilar una muestra es:

$$C = n \cdot c_1 + n \cdot m \cdot c_2, \qquad [10]$$

donde c_1 es el costo (tiempo) esperado para la medición de una unidad primaria y c_2 es el costo (tiempo) esperado para la medición de los segmentos restantes de un path.

La solución práctica que se propone se basa en una aproximación gráfica. La solución puede ser implementada en árboles que han sido completamente medidos. Requiere conocer la varianza entre unidades primarias y la varianza de los paths que se generan a partir de las unidades primarias (ver [5]), como también los costos respectivos (ver [10]) obtenidos de diferentes combinaciones de n, m en el rango de tamaños n posibles.

El primer paso es construir un gráfico con el error estándar de estimación (ordenada) y el tamaño muestral n (absisa). Allí se construyen las curvas de error estándar para diferentes tamaños secundarios de muestra (esto es m = 1, m = 2, ...,). A continuación se trazan curvas de costo. Para ello se determina el tamaño m necesario para lograr un costo fijo C para cada tamaño n entre 2 y el máximo tamaño n posible (ver restricciones de tamaño más atrás), esto es mediante:

$$m = \frac{C - n \cdot c_1}{n \cdot c_2}. \qquad [11]$$

Este tamaño de muestra se utiliza para superponer la curva de costo C en el gráfico de error estándar. La posición exacta de la curva de costo

C para cada tamaño de muestra *n* se determina calculando el error estándar que proveería una muestra de tamaño *n*, *m* calculado; para ello se requiere contar con las varianzas entre y dentro de unidades primarias. El proceso se repite para diferentes valores *C*. Al final se dispone de un gráfico que entrega el costo de diferentes configuraciones muestrales (*n*, *m*) junto con la precisión que otorgan, de donde el usuario puede elegir la configuración más adecuada a sus intereses.

El proceso de optimización en el muestreo multietápico se concentra en la determinación del costo mínimo para lograr una precisión específica o en la determinación de la precisión máxima dado un costo total prefijado. El primer caso se soluciona aquí trazando una línea horizontal en el nivel de precisión deseado, seleccionando la curva menor y encontrando la combinación n, m más cercana. El segundo caso se soluciona desplazándose a lo largo de la curva de costo deseado hasta detectar el punto de máxima precisión e identificando la combinación n, m más cercana.

El método expuesto se incorporó a BRANCH (7). La implementación del método se ilustra a continuación en dos árboles (*Picea abies* (L.) Karst.), los cuales fueron medidos en su totalidad. Se asumió que todas las ramas principales constituyen un solo nudo, es decir, no se consideran los segmentos del fuste como segmentos elegibles. Los árboles varían fuertemente en la proporción en que las unidades primarias contribuyen a la varianza del estimador (7). En el análisis se consideran tres niveles de costo total y se varía la relación de costos entre unidades primarias y resto de path.

El costo o tiempo de medición incluye el tiempo requerido en cada nodo para contabilizar los segmentos, medir la variable auxiliar, seleccionar segmentos, y marcar y obtener la variable de interés en los segmentos seleccionados. Así, diferentes niveles de costo pueden generarse para las unidades primarias (c_1) y para el resto del path (c_2). El costo por unidad primaria es bajo en árboles pequeños, con pocas unidades primarias y con unidades primarias sin variable de interés; en cambio, c_1 es alto en árboles grandes, con numerosas unidades primarias, ya sea que éstas se midan a lo largo del fuste o que se corten y ordenen para posteriormente medirlas. El costo c_2 tiene relación con el tamaño de los path; mientras más largo el path, más alto es c_2. En resumen, variados niveles de costo pueden obtenerse de un árbol en particular, dependiendo de la estructura definida para su muestreo. Ello justifica la incorporación de diferentes relaciones de costo c_1, c_2.

RESULTADOS Y DISCUSION

Considerando un tamaño de muestra $n = 1$ y $m = 1$, y definiendo $\sigma_1^2 = \sum_{i=1}^{N} q_i \left(\frac{F_i}{q_i} - \sum_{i=1}^{N} F_i \right)^2$ como la variabilidad básica entre unidades muestrales y $\sigma_{resto}^2 = \sum_{i=1}^{N} \frac{1}{q_i} \sum_{j=1}^{M_{Path.i}} Q_{R_{1ij}} \left(\hat{F}_{1ij} - F_i \right)^2$ la variabilidad básica de los paths que se inician en las unidades muestrales primarias, [2''] se transforma en $Var\hat{F} = \sigma_1^2 + \sigma_{resto}^2$; así, $Pvp = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_{resto}^2}$ entrega la proporción de variabilidad atribuible a las unidades primarias. Esta proporción es, a primera vista, determinante del éxito probable de la disminución del error estándar al aumentar el tamaño de la muestra primaria. Sin embargo, los resultados obtenidos con la utilización del RBS swor_swr sólo presentan una leve evidencia en esta dirección. Por ejemplo, al aumentar el tamaño de muestra primario de $n = 2$ a $n = 9$, la disminución en el error estándar con $m = 1$ y $m = 8$ es, respectivamente, 57,4% y 61,7% con *Pvp* alto (cuadro 1), y es sólo levemente menor con *Pvp* bajo (54,1% y 58,8%, respectivamente; cuadro 2). Esto evidencia que el aumento de *n* no disminuye el error estándar tan fuertemente como podría esperarse en el árbol con alta *Pvp*.

La *Pvp* no es sólo un indicador de la variabilidad entre unidades primarias sino también un indicador de la variabilidad dentro de esas unidades, esto es de la variabilidad de los paths que se inician en las unidades primarias. Así, una baja *Pvp* es a su vez un indicador de una alta variabilidad dentro de las unidades primarias. Así, la disminución del error estándar está estrechamente ligada al aumento de *m*, siendo ésta más fuerte mientras más baja es *Pvp*. Esto es evidente al comparar los resultados obtenidos (cuadros 1 y 2; figura 1). Al aumentar de $m = 1$ a $m = 8$, con $n = 2$ y $n = 9$, la disminución del error estándar es de 26,4% y 33,7%, respectivamente, en el árbol con *Pvp* alta. En cambio, en el árbol con *Pvp*

CUADRO 1

Error estándar (%) (ver relación [5]) para diferentes combinaciones de tamaños muestrales (*n*, *m*) para el árbol 1 (*Pvp* = 0,49; tamaño muestral primario máximo = 10).
Standard error (%) (see relation [5]) for different combinations of sample sizes (*n*, *m*) for tree 1; maximum primary sample size = 10).

n/*m*	1	2	3	4	5	6	7	8
2	39,0	33,5	31,5	30,4	29,7	29,2	28,9	28,7
3	31,5	26,9	25,2	24,3	23,7	23,3	23,1	22,8
4	26,9	22,9	21,3	20,5	20,1	19,7	19,5	19,3
5	23,7	20,1	18,7	17,9	17,5	17,2	17,0	16,8
6	21,4	17,9	16,7	16,0	15,6	15,3	15,1	14,9
7	19,5	16,3	15,1	14,4	14,0	13,7	13,5	13,4
8	17,9	14,9	13,7	13,1	12,7	12,4	12,2	12,1
9	16,6	13,7	12,6	12,0	11,6	11,3	11,1	11,0
10	15,5	12,7	11,6	11,0	10,6	10,4	10,2	10,0

CUADRO 2

Error estándar (%) (ver relación [5]) para diferentes combinaciones de tamaños muestrales (*n*, *m*) para el árbol 2 (*Pvp* = 0,09 tamaño muestral primario máximo = 9).
Standard error (%) (see relation [5]) for different combinations of sample sizes (*n*, *m*) for tree 2 (*Pvp = 0.09;* maximum primary sample size = 9).

n/*m*	1	2	3	4	5	6	7	8
2	43,1	31,8	26,9	24,1	22,3	21,0	20,0	19,2
3	35,1	25,8	21,8	19,5	18,0	16,9	16,1	15,5
4	30,3	22,2	18,7	16,7	15,4	14,5	13,7	13,2
5	27,0	19,7	16,6	14,8	13,6	12,7	12,1	11,6
6	24,6	17,9	15,0	13,4	12,3	11,5	10,8	10,4
7	22,7	16,5	13,8	12,2	11,2	10,4	9,9	9,4
8	21,1	15,3	12,8	11,3	10,3	9,6	9,0	8,6
9	19,8	14,3	11,9	10,5	9,5	8,8	8,3	7,9

baja, la disminución es bastante más fuerte (55,4% y 60,1%, respectivamente).

El logro de un tamaño primario óptimo dentro del rango de tamaños primarios posibles depende de los costos y de las varianzas. En general, mientras menor es la variabilidad de las unidades primarias, la máxima precisión para un nivel de costo total dado se logra con tamaños de muestra primaria menores. Como es lógico, con baja disponibilidad de recursos, el nivel de precisión lo-

árbol 1 árbol 2

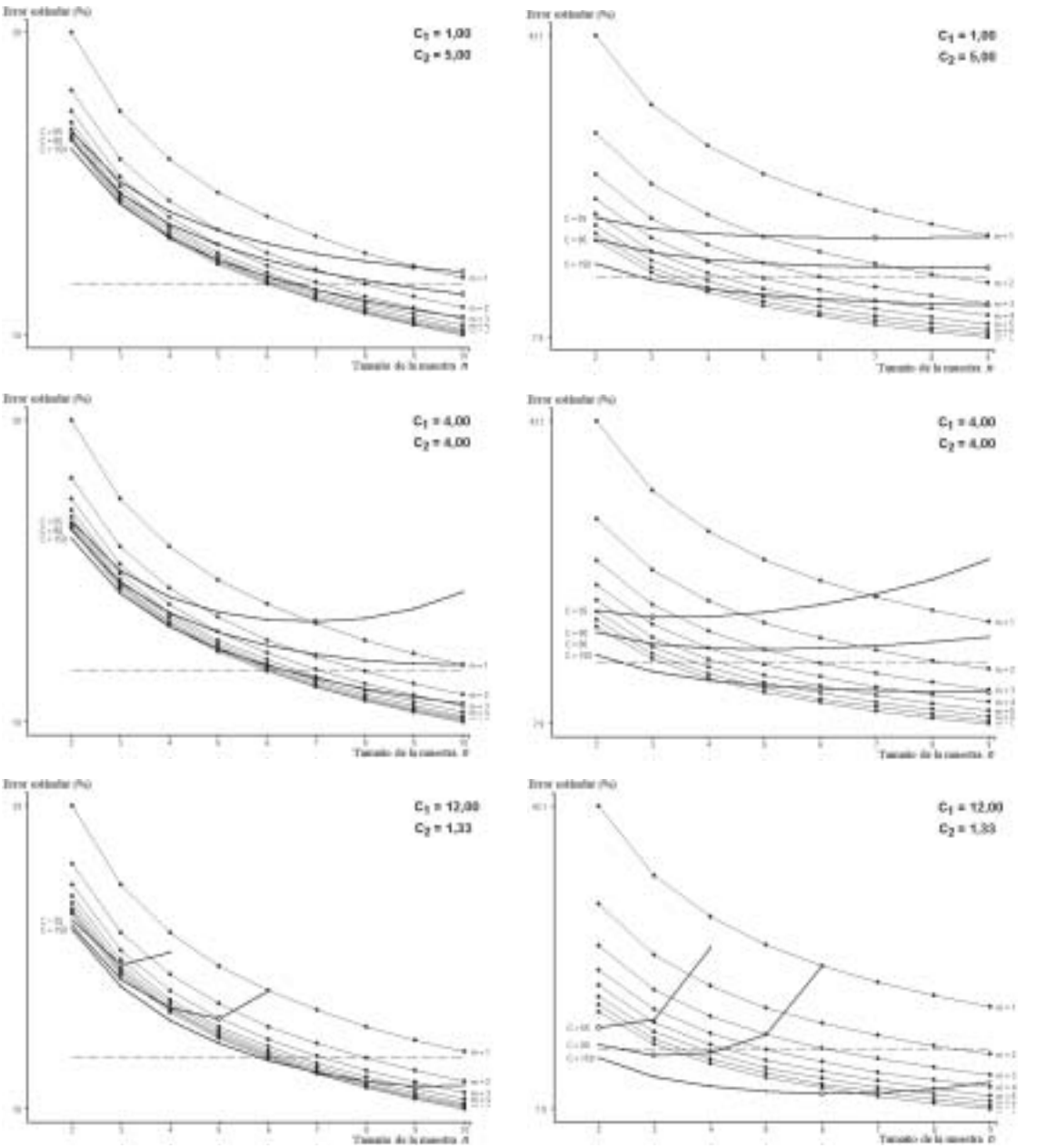

Figura 2. Error estándar (líneas delgadas) para diferentes combinaciones de tamaños de muestra (*n*, *m*) y curvas de costos totales de muestreo para tres niveles de costo total (líneas gruesas) y tres relaciones de costo esperado (c_1, c_2) en dos árboles con diferente proporción de varianza entre unidades muestrales (izquierda *Pvp* = 0,49; derecha *Pvp* = 0,09). El círculo en cada curva de costos señala el menor error estándar en el rango de tamaños de muestra analizado. La línea segmentada representa un error estándar de 15%.

Standard error (thin lines) for different combinations of sample sizes (*n*, *m*) and curves of total costs of sampling for three levels of total cost (heavy lines) and three relations of expected cost (c_1, c_2) in two trees with different variance proportions between sampling units (left *Pvp* = 0.49, right *Pvp* = 0.09). The circle in each curve of costs indicates the smaller standard error in the rank of analyzed sample sizes. The segmented line represents a standard error of 15%.

grado es a su vez bajo, pero es más fácil detectar el nivel de máxima precisión dado un nivel de gasto fijo. A menor nivel de gasto, menor es el tamaño de muestra *n* que minimiza la varianza.

El tamaño óptimo de muestra primaria tiende a ser grande y a exceder el máximo posible cuando el costo por unidad primaria es bajo en comparación al costo de medir el resto del path. Esto es más evidente en árboles con una alta proporción de varianza atribuible a las unidades primarias. A medida que aumenta el costo de las unidades primarias en relación al costo de medición del resto del path (cuadros 3, 4 y 5; figura 2) la máxima precisión se logra con menor tamaño de muestra primaria y mayor tamaño de muestra secundaria.

El punto de mínimo costo para un determinado error estándar se desplaza dependiendo de la relación de varianzas y costos. A mayor costo

CUADRO 3

Costo total esperado para diferentes combinaciones de tamaños muestrales (*n*, *m*) (ver relación [10]) con c_1 = 1,00 y c_2 = 5,00. (Nota: la tabla es válida para ambos árboles en análisis; para el árbol 2 observar sólo hasta *n* = 9).

Expected total cost for different combinations of sample sizes (*n*, *m*) (see relation [10]) with c_1 = 1.00 and c_2 = 5.00 (Note: the table is valid for both trees in the analysis; for tree 2 consider only up to *n* = 9)

n/*m*	1	2	3	4	5	6	7	8
2	12,0	22,0	32,0	42,0	52,0	62,0	72,0	82,0
3	18,0	33,0	48,0	63,0	78,0	93,0	108,0	123,0
4	24,0	44,0	64,0	84,0	104,0	124,0	144,0	164,0
5	30,0	55,0	80,0	105,0	130,0	155,0	180,0	205,0
6	36,0	66,0	96,0	126,0	156,0	186,0	216,0	246,0
7	42,0	77,0	112,0	147,0	182,0	217,0	252,0	287,0
8	48,0	88,0	128,0	168,0	208,0	248,0	288,0	328,0
9	54,0	99,0	144,0	189,0	234,0	279,0	324,0	369,0
10	60,0	110,0	160,0	210,0	260,0	310,0	360,0	410,0

CUADRO 4

Costo total esperado para diferentes combinaciones de tamaños muestrales (*n*, *m*) (ver relación [10]) con c_1 = 4,00 y c_2 = 4,00. (Nota: la tabla es válida para ambos árboles en análisis; para el árbol 2 observar sólo hasta *n* = 9).

Expected total cost for different combinations of sample sizes (*n*, *m*) (see relation [10]) with c_1 = 4.00 and c_2 = 4.00. (Note: the table is valid for both trees in the analysis; for tree 2 consider only up to *n* = 9)

n/*m*	1	2	3	4	5	6	7	8
2	16,0	24,0	32,0	40,0	48,0	56,0	64,0	72,0
3	24,0	36,0	48,0	60,0	72,0	84,0	96,0	108,0
4	32,0	48,0	64,0	80,0	96,0	112,0	128,0	144,0
5	40,0	60,0	80,0	100,0	120,0	140,0	160,0	180,0
6	48,0	72,0	96,0	120,0	144,0	168,0	192,0	216,0
7	56,0	84,0	112,0	140,0	168,0	196,0	224,0	252,0
8	64,0	96,0	128,0	160,0	192,0	224,0	256,0	288,0
9	72,0	108,0	144,0	180,0	216,0	252,0	288,0	324,0
10	80,0	120,0	160,0	200,0	240,0	280,0	320,0	360,0

CUADRO 5

Costo total esperado para diferentes combinaciones de tamaños muestrales (n, m) (ver relación [10]) con c_1 = 12,00 y c_2 = 1,33. (Nota: la tabla es válida para ambos árboles en análisis; para el árbol 2 observar sólo hasta n = 9).

Expected total cost for different combinations of sample sizes (n, m) (see relation [10]) with c_1 = 12.00 and c_2 = 1.33. (Note: the table is valid for both trees in the analysis; for tree 2 consider only up to n = 9)

n/m	1	2	3	4	5	6	7	8
2	26,66	29,32	31,98	34,64	37,30	39,96	42,62	45,28
3	39,99	43,98	47,97	51,96	55,95	59,94	63,93	67,92
4	52,32	58,64	63,96	69,28	74,60	79,92	85,24	90,56
5	66,65	73,30	79,95	86,60	93,25	99,90	106,55	113,20
6	79,98	87,96	95,94	103,92	111,90	119,88	127,86	135,84
7	93,31	102,62	111,93	121,24	130,55	139,86	149,17	158,48
8	106,64	117,28	127,92	138,56	149,20	159,84	170,48	181,12
9	119,97	131,94	143,91	155,88	167,85	179,82	191,79	203,76
10	133,30	146,60	159,90	173,20	186,50	199,80	213,10	226,40

CUADRO 6

Tamaño máximo de muestra m para diferentes niveles de costo total (C), tamaños de muestra primario (n) y distintas relaciones de costo esperado c_1, c_2 (ver relación [11]). (Nota: la tabla es válida para ambos árboles en análisis; para el árbol 2 observar sólo hasta n = 9).

Maximum sample size (m) for different levels of total cost (c), primary sample sizes (n) and different relations of expected costo c_1, c_2 (see relation [11]). (Note: the table is valid for both trees in the analysis; for tree 2 consider only up to n = 9)

	c_1 = 1,00; c_2 = 5,00			c_1 = 4,00; c_2 = 4,00			c_1 = 12,00; c_2 = 1,33		
n/C	55	80	150	55	80	150	55	80	150
2	5,30	7,80	14,80	5,88	9,00	17,75	11,65	21,05	47,37
3	3,47	5,13	9,80	3,58	5,67	11,50	4,76	11,03	28,57
4	2,55	3,80	7,30	2,44	4,00	8,38	1,32	6,02	19,17
5	2,00	3,00	5,80	1,75	3,00	6,50		3,01	13,53
6	1,63	2,47	4,80	1,29	2,33	5,25		1,00	9,77
7	1,37	2,09	4,09	0,96	1,86	4,36			7,09
8	1,18	1,80	3,55	0,72	1,50	3,69			5,08
9	1,02	1,58	3,13	0,53	1,22	3,17			3,51
10	0,90	1,40	2,80	0,38	1,00	2,75			2,26

relativo de unidades primarias y a mayor proporción de varianza dentro de las unidades muestrales primarias, el costo mínimo para un nivel de precisión específico se logra con menor *n* y mayor *m*. Esto porque tanto el mayor costo por unidad primaria, lo que determina un gran cambio en *m* por cada decremento unitario en *n* (cuadro 6), como el gran efecto de *m* sobre el error estándar, favorecen la captura de muestras con *m* alto y *n*.

CONCLUSION

El método propuesto proporciona flexibilidad suficiente para ser utilizado en la determinación de tamaños de muestra al aplicar el muestreo RBS sworswr. La identificación de la combinación de tamaños n, m que provee el costo mínimo para lograr una precisión específica o la precisión máxima dado un costo total prefijado se realiza fácilmente a partir del gráfico generado. La investigación en esta línea debe concentrarse en la extensión del método a árboles en los que se ha obtenido una muestra.

BIBLIOGRAFIA

(1) VALENTINE, H.T., L.M. TRITTON, G.M.FURNIVAL. Subsampling trees for biomass, volume, or mineral content. *For. Sci.* 1984, vol. 30, p. 673-681.
(2) JESSEN, R.J. Determining the fruit count on a tree by randomized branch sampling. *Biometrics*, 1955, vol. 11, p. 99-109.
(3) VALENTINE, H.T., Jr. V.C. BALDWIN, T.G. GREGOIRE, H.E. BURKHART. Surrogates for foliar dry matter in loblolly pine. *For. Sci.* 1994, vol. 40, p. 576-585.
(4) VALENTINE, H.T., S.J. HILTON. Sampling oak foliage by the randomized-branch method. *Can. J. For. Res.*, 1977, vol. 7, p. 295-298.
(5) WILLIAMS, R.A. Use of randomized branch and importance sampling to estimate loblolly pine biomass. *South. J. Appl. For.*, 1989, vol. 13, p. 181-184.
(6) GROSENBAUGH, L.R. The gains from sample-tree selection with unequal probabilities. *Journal of Forestry*, 1967, vol. 65, p. 202-206.
(7) CANCINO, J. Analyse und praktische Umsetzung unterschiedlicher Methoden des Randomized Branch Sampling. Diss. Fakultät für Forstwissenschaften und Waldökologie der Georg-August-Universität Göttingen. 191 S. http.//webdoc.sub.gwdg.de/diss/2003/cancino/index.html.
(8) GREGOIRE, T.G., H.T. VALENTINE, G.M. FURNIVAL. Sampling methods to estimate foliage and other characteristics of individual trees. *Ecology*, 1995, vol. 76, p. 1.181-1.194.
(9) PARRESOL, B.R. Assessing tree and stand biomass. A review with examples and critical comparisons-For. Sci., 1999, vol. 45, p. 573-593.
(10) GOOD, M., M. PATERSON, C. BRACK, K. MENGERSEN. Estimating tree component biomass using variable probability sampling methods. *Journal of Agricultural, Biological and Environmental Statistics*, 2001, vol. 6, p. 258-267.
(11) SNOWDON, P., J. RAISON, H. KEITH, K. MONTAGU, H. BI, P. RITSON, P. GRIERSON, M. ADAMS, W. BURROWS, D. EAMUS. Protocol for sampling tree and stand biomass. National Carbon Accounting System, 2001, Technical Report Nº 31. 114 p.
(12) SABOROWSKI, J., D. GAFFREY, D. RBS. Ein mehrstufiges Inventurverfahren zur Schätzung von Baummerkmalen. II. Modifizierte RBS-Verfahren. AFJZ, 1999, vol. 170, p. 223-227.
(13) COCHRAN, W.G. *Sampling techniques.* Wiley, New York., 1977, 428 p.
(14) SAMPFORD, M. R. On sampling without replacement with unequal probabilities of selection. *Biometrika*, 1967, vol. 54, p. 499-513.
(15) SABOROWSKI, J. Schätzung von Varianzen und Konfidenzintervallen aus mehrstufigen Stichproben. Schriften aus der Forstlichen Fakultät der Universität Göttingen und der Niedersächsischen Forstlichen Versuchsanstalt, Bd., 1990, 99 p.
(16) HORVITZ, D.G., D.J. THOMPSON. A generalisation of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 1952, vol. 47, p. 663-685.

Recibido: 10.11.04.
Aceptado: 24.03.05.