

# Cut-off Importance Sampling of Bole Volume

Andrew P. Robinson, Timothy G. Gregoire and Harry T. Valentine

---

**Robinson, A.P., Gregoire, T.G. & Valentine, H.T.** 1997. Cut-off importance sampling of bole volume. *Silva Fennica* 31(2): 153–160.

Cut-off importance sampling (CIS) is introduced as a means of sampling individual trees for the purpose of estimating bole volume. The novel feature of this variant of importance sampling is the establishment on the bole of a cut-off height,  $H_C$ , above which sampling is precluded. An estimator of bole volume between predetermined heights  $H_L$  and  $H_U > H_C$  is proposed, and its design-based bias and mean square error are derived. In an application of CIS as the second stage of a two-stage sample to estimate aggregate bole volume, the gain in precision realized from CIS more than offset its bias when compared to the precision of importance sampling when  $H_C = H_U$ .

**Keywords** mean-square error, Monte Carlo methods, two-stage sampling

**Authors' addresses** *Robinson* (arobins@forestry.umn.edu), Department of Forest Resources, University of Minnesota, St. Paul, MN 55108-6112, USA; *Gregoire* (tgg@vt.edu), College of Forestry and Wildlife Resources, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061-0324, USA; *Valentine*, USDA Forest Service, Durham, NH 03824-0640, USA

**Accepted** 18 April 1997

---

## 1 Introduction

Importance sampling (IS) has received considerable attention in both the forestry and statistics literature as a method of estimating bole volume (e.g., Gregoire et al. 1986, 1993, 1995; Wiant et al. 1989; Van Deusen 1990; Valentine et al. 1992; Schreuder et al. 1993; Robinson and Wood 1994). Application of the method ordinarily requires the measurement of diameters or cross-sectional areas of the bole at heights selected at random. The ease with which a bole measurement can be made on a standing tree depends upon the select-

ed height and whether it is within the crown of the tree. High measurement points may be obscured from view or difficult to locate exactly (Wood and Wiant 1992).

Särndal et al. (1992) discuss the possibility of cut-off sampling, in which a part of the target population is deliberately excluded from the sampling frame. This is presented as a compromise between probability sampling and nonprobabilistic selection and it leads to biased estimators. It is considered reasonable when it would cost too much to construct and maintain a complete frame, and the bias is not expected to be very great.

We suggest “cut-off importance sampling” (CIS) as a means to avoid measurements in the uppermost section of a bole. The technique takes its name from the fact that the sampling is restricted to the region of the bole beneath a cut-off height or “cut-off.” The cut-off is determined for each bole prior to sampling. The restriction of the IS to beneath the cut-off tends to reduce the sampling variance. Of course, there is a downside to CIS: the estimate of the volume above the cut-off is biased. However, our studies indicate that, for small sample sizes, cut-off importance sampling yields an overall reduction in mean-square error.

## 2 Method

The estimation of bole volume of a standing tree by IS ordinarily requires a measurement of the height ( $H$ ) of the tree at the outset. The sampling also requires auxiliary information in the form of an integrable “proxy taper function” (ptf). The ptf defines the cross-sectional area of a “proxy bole” and predicts the cross-sectional area of the bole of interest at any height  $0 \leq h \leq H$ . The ptf is used to construct a probability density function for  $h$  from which the sample heights are selected at random (see, e.g., Gregoire et al. 1986). The ptf need not be specifically fitted to the species being sampled; a simple generic taper function will suffice. However, a very accurate ptf – in the sense that the taper of the proxy bole is nearly proportional to the taper of the bole of interest – will afford very efficient IS.

### Preliminaries

Let  $A(h)$  denote cross-sectional area of the bole of interest at height  $h$  and let  $A_p(h)$  denote the cross-sectional area of the proxy bole defined by the ptf at height  $h$ . The volume of the bole between the limits of interest,  $H_L$  and  $H_U$ , is denoted by  $V(H_L, H_U)$ , therefore,

$$V(H_L, H_U) = \int_{H_L}^{H_U} A(h) dh$$

Similarly, the volume of the proxy bole defined by the ptf between the limits  $H_L$  and  $H_U$  is denoted by  $V_p(H_L, H_U)$ , therefore,

$$V_p(H_L, H_U) = \int_{H_L}^{H_U} A_p(h) dh$$

The height of the cut-off is denoted by  $H_C$  and the volumes of the bole and the proxy bole between  $H_L$  and  $H_C$  are denoted, respectively, by  $V(H_L, H_C)$  and  $V_p(H_L, H_C)$ . Finally, let  $\hat{V}(H_L, H_U)$  denote the usual unbiased IS estimator of  $V(H_L, H_U)$ , i.e.,

$$\hat{V}(H_L, H_U) = \frac{1}{m} \sum_{i=1}^m \frac{A(\phi_i)}{f(\phi_i)} \tag{1a}$$

$$= \frac{V_p(H_L, H_U)}{m} \sum_{i=1}^m \frac{A(\phi_i)}{A_p(\phi_i)} \tag{1b}$$

where  $h = \phi_i$  ( $i = 1, \dots, m$ ) is selected at random from the probability density function:

$$f(h) = \begin{cases} A_p(h) / V_p(H_L, H_U), & \text{if } H_L \leq h \leq H_U \\ 0, & \text{otherwise} \end{cases}$$

The variance of  $\hat{V}(H_L, H_U)$  is:

$$\text{Var}(\hat{V}(H_L, H_U)) = \frac{1}{m} \int_{H_L}^{H_U} f(h) \cdot \left[ \frac{A(h)}{f(h)} - V(H_L, H_U) \right]^2 dh \tag{2}$$

### Cut-off Sampling

Under the cut-off scheme, sampling is restricted between  $H_L$  and  $H_C$ , where  $H_C < H_U$ . We define the cut-off probability density function,  $g(h)$ , in the usual manner, i.e.,

$$g(h) = \begin{cases} A_p(h) / V_p(H_L, H_C), & \text{if } H_L \leq h \leq H_C \\ 0, & \text{otherwise} \end{cases}$$

If heights  $h = \theta_i$  ( $i = 1, \dots, m$ ) are selected at random from  $g(h)$ , then the usual unbiased estimator of the volume from  $H_L$  to the cut-off height,  $H_C$ , is:

$$\hat{V}(H_L, H_C) = \frac{1}{m} \sum_{i=1}^m \frac{A(\theta_i)}{g(\theta_i)} = \frac{V_p(H_L, H_C)}{m} \sum_{i=1}^m \frac{A(\theta_i)}{A_p(\theta_i)}$$

Presumably the volume of the bole from  $H_L$  to  $H_U$  remains the parameter of interest. As an estimator of  $V(H_L, H_U)$  under CIS we propose a ratio adjustment of the unbiased estimator  $\hat{V}(H_L, H_C)$ , namely

$$\tilde{V}(H_L, H_U) = \frac{V_p(H_L, H_U)}{V_p(H_L, H_C)} \hat{V}(H_L, H_C) \tag{3a}$$

The analogy of  $\tilde{V}(H_L, H_U)$  to  $\hat{V}(H_L, H_U)$  is direct upon defining

$$f^*(h) = \begin{cases} A_p(h) / V_p(H_L, H_U), & \text{if } H_L \leq h \leq H_C \\ 0, & \text{otherwise} \end{cases}$$

and re-expressing  $\tilde{V}(H_L, H_U)$  as

$$\tilde{V}(H_L, H_U) = \frac{1}{m} \sum_{i=1}^m \frac{A(\theta_i)}{f^*(\theta_i)} \tag{3b}$$

The essential difference between (1a) and (3b) is that  $f^*(\cdot)$  in (3b) is not a probability density, as it does not integrate to unity. This estimator of  $V(H_L, H_U)$  is biased, in general:

$$\begin{aligned} B &= E[\tilde{V}(H_L, H_U)] - V(H_L, H_U) \\ &= \int_{H_L}^{H_C} \frac{A(h)}{f^*(h)} g(h) dh - V(H_L, H_U) \\ &= \frac{V_p(H_L, H_U)}{V_p(H_L, H_C)} V(H_L, H_C) - V(H_L, H_U) \\ &= V(H_L, H_C) \left[ \frac{V_p(H_L, H_U)}{V_p(H_L, H_C)} - \frac{V(H_L, H_U)}{V(H_L, H_C)} \right] \end{aligned} \tag{4}$$

It is evident from the expression in square brackets that  $\tilde{V}(H_L, H_U)$  unbiasedly estimates  $V(H_L, H_U)$  only if the proportions of the total volume above and below the cut-off are identical for both the bole of interest and the proxy bole. The CIS estimator of  $V(H_L, H_U)$  can be partitioned into two components:

$\hat{V}(H_L, H_C)$ , which is unbiased for  $V(H_L, H_C)$ ; and  $\tilde{V}(H_C, H_U) = \tilde{V}(H_L, H_U) - \hat{V}(H_L, H_C)$ , which subsumes all of the bias of  $\tilde{V}(H_L, H_U)$ . Ordinarily, however, the magnitude of  $V(H_L, H_C)$  will be far greater than that of  $V(H_C, H_U)$ , and hence there is a priori reason to believe that the bias of  $\tilde{V}(H_L, H_U)$  will be small relative to  $V(H_L, H_U)$ .

The mean-square error of  $\tilde{V}(H_L, H_U)$  as an estimator of  $V(H_L, H_U)$  is:

$$\begin{aligned} \text{MSE}(\tilde{V}(H_L, H_U)) &= \frac{1}{m} \int_{H_L}^{H_C} g(h) \left[ \frac{A(h)}{f^*(h)} - V(H_L, H_U) \right]^2 dh \end{aligned} \tag{5}$$

### 3 Test

The statistical performance of  $\tilde{V}(H_L, H_U)$  under CIS was compared to that of  $\hat{V}(H_L, H_U)$  under IS with the aid of detailed stem measurements on five species of trees: loblolly pine (*Pinus taeda* L.), southern red oak (*Quercus falcata* Michx L.), slash pine (*Pinus elliottii* Engelm.), sweetgum (*Liquidambar styraciflua* L.), and white oak (*Quercus alba* L.). Information summarizing tree sizes is displayed in Table 1.

Each tree in the database had been felled and cut into roughly 1 m sections. The cross-sectional area at the base of each section was calculated from a measurement of outside-bark diameter. We used a cubic spline to interpolate between the successive cross-sectional areas of each bole. Thus the cubic spline defined  $A(h)$  for  $H_L \leq h \leq H_U$  and its integral gave  $V(H_L, H_U)$ . Any deviation between this determination of bole volume and that obtainable from gravimetric techniques was assumed to be inconsequential. An advantage to determining bole volume as the integral of bole cross-sectional area is the ability to compute volume to any stipulated upper-bole diameter. Proxy boles were defined by the following ptf:

$$A_p(h) = A(H_B) \left[ \frac{H-h}{H-H_B} \right]^{4/c}$$

where  $H_B$  denotes breast height. Following Gre-

**Table 1.** Summary information for the five sets of tree measurements used in the study of  $\tilde{V}(H_L, H_U)$  as an estimator of bole volume following CIS.

Species	N	Dbh			Total stem height			Total stem volume		
		Min	Mean	Max	Min	Mean	Max	Min	Mean	Max
		cm			m			m <sup>3</sup>		
Loblolly pine	89	2.0	13.5	31.2	2.3	9.8	18.7	.003	.108	.563
Red oak	50	16.0	43.7	88.1	19.4	27.1	35.4	.199	2.138	7.733
Slash pine	78	1.8	12.2	28.2	1.7	8.8	17.7	.005	.097	5.207
Sweetgum	39	14.7	39.9	75.4	18.0	26.9	32.7	.169	1.866	5.890
White oak	38	14.5	40.4	78.3	14.0	23.8	32.2	.121	1.631	5.301

goire et al. (1995), we let the shape parameter,  $c$ , take values of 3 or 4.

For the sake of illustrating the performance of the suggested sampling strategy, a two-stage design was implemented to estimate the aggregate bole volume in each population. The first stage consisted of the selection of trees from the population by list sampling with replacement using selection probabilities proportional to the integrated ptf, namely  $V_p(H_L, H_U)$ .

The second stage consisted of the independent selection of  $m$  sampling heights on each tree chosen in stage 1 by both IS and CIS. In all cases the lower limit of integration,  $H_L$ , matched the stump height of the tree. The upper limit,  $H_U$ , was alternately set at total tree height, the height to a upper-bole diameter of 5 cm (2 inches), or of 10 cm (4 inches), in order to compare the effect that varying  $H_U$  has on the performances of the estimator of aggregate volume.

The target parameter estimated by the two-stage sampling was  $V = \sum_{k=1}^N V_k$  where  $V_k$  denotes the volume,  $V(H_L, H_U)$  of the  $k$ th of  $N$  boles. Let  $\hat{V}_{2S}$  or  $\tilde{V}_{2S}$ , respectively, denote the estimator of  $V$  where IS or CIS is the second-stage method. In the first case, the estimator of  $V$  is:

$$\hat{V}_{2S} = \frac{1}{n} \sum_{k=1}^n \frac{\hat{V}_k}{P_k}$$

where  $\hat{V}_k = \hat{V}(H_L, H_U)$  for the  $k$ th bole (eqn (1)) and  $P_k$  is the first-stage selection probability of the  $k$ th bole from the population of  $N$  boles. Its variance is:

$$\text{Var}(\hat{V}_{2S}) = \frac{1}{n} \sum_{k=1}^N \left[ P_k \left( \frac{V_k}{P_k} - V \right)^2 + \frac{\text{Var}(\hat{V}_k)}{P_k} \right] \quad (6)$$

where  $\text{Var}(\hat{V}_k)$  is the variance of  $\hat{V}(H_L, H_U)$  for the  $k$ th bole (eqn (2)). When CIS is used in the second stage, the estimator of  $V$  is:

$$\tilde{V}_{2S} = \frac{1}{n} \sum_{k=1}^n \frac{\tilde{V}_k}{P_k} \quad (7)$$

The bias of  $\tilde{V}_{2S}$  is

$$B(\tilde{V}_{2S}) = \sum_{k=1}^N \frac{B_k}{P_k} \quad (8)$$

and its mean-square error is

$$\text{MSE}(\tilde{V}_{2S}) = \frac{1}{n} \sum_{k=1}^N \left[ P_k \left( \frac{V_k}{P_k} - V \right)^2 + \frac{\text{MSE}(\tilde{V}_k)}{P_k} \right] \quad (9)$$

where  $\tilde{V}_k$  and  $\text{MSE}(\tilde{V}_k)$  obtain from eqns (3) and (4), respectively.

### 4 Results

The results in Table 2 pertain to the case where  $n = 1$  tree was selected in the first stage and  $m = 2$  heights were selected from either  $f(h)$  (for IS) or  $g(h)$  (for CIS) in the second stage. The target volume,  $V$ , was aggregate bole volume to the height of a 5 cm (2 inch) upper-bole diame-



**Table 2.** Summary of two-stage samplings of aggregate bole volume. The first- and second-stage sample sizes, respectively, were  $n = 1$  and  $m = 2$ . Bole volume to a 5 cm top diameter was estimated. SE signifies the standard error of  $\hat{V}_{2S}$ . RMSE and Bias signify the root mean-square error and bias of  $\tilde{V}_{2S}$ , respectively. The cut-offs were 60 % or 80 % of tree height. SE, RMSE, and Bias are presented as percentages of the true aggregate volume to a 5 cm top diameter.

Species	$c^\dagger$	$\hat{V}_{2S}$	$\tilde{V}_{2S}$ (80 %)		$\tilde{V}_{2S}$ (60 %)	
		SE	RMSE	Bias	RMSE	Bias
Loblolly pine	3	12.9	12.7	0.4	11.8	2.6
	4	18.3	17.8	1.0	15.9	5.8
Red oak	3	20.3	19.6	1.2	17.8	4.9
	4	23.2	21.6	2.5	18.8	8.8
Slash pine	3	12.7	12.7	0.3	12.5	1.6
	4	15.7	15.3	0.9	14.5	4.0
Sweetgum	3	17.9	16.9	1.2	15.6	3.5
	4	19.8	18.0	2.4	16.3	6.7
White oak	3	21.7	21.0	1.2	19.4	5.5
	4	26.3	24.9	2.6	22.1	10.2

$^\dagger$  Shape parameter of the the proxy taper function.

ter. Therefore  $H_U$  varied in each tree. Standard errors (SE) for  $\hat{V}_{2S}$  were calculated exactly from (6), because we knew the actual volume,  $V_k$ , and first-stage selection probability,  $P_k$ , of each tree in the population, and we could integrate the cubic-spline profile of each tree to evaluate  $\text{Var}(\hat{V}_k)$  for each tree, as well. Bias and root-mean-square errors (RMSE) of  $\tilde{V}_{2S}$  were calculated exactly using (5) and (9). All results in Table 2 are expressed as a percentage of  $V$ . We also calculated errors for samplings to the height of an upper diameter of 10 cm (4 inches); the net effect was reduced bias and RMSE for all cases compared to the results in Table 2. The bias percentages reported in Table 2 are invariant to the size of both stages of sampling. The SE and RMSE results shown in the table can be prorated to first-stage samples of size  $n > 1$  by dividing by  $\sqrt{n}$ .

For all five tree species, the RMSE of the estimate uniformly decreased as the cut-off lowered. The bias increased at varying rates, and we note that it was worse for those species for which the ptf was a poor fit (Table 2), specifically, the overall cost was higher for hardwoods than soft-

woods, because the ptf was a better fit to the latter.

Profiles are presented for slash pine and white oak in Fig. 1. Note that the RMSE of  $\tilde{V}_{2S}$  at  $H_C = H_U$  is equivalent to the SE of  $\hat{V}_{2S}$ , so to compare the performance of CIS at any particular point with IS, one needs to compare it with CIS at 100 % on the same graph. The graphs both clearly show that, as the cut-off descends the bole, the RMSE remains stable or actually decreases, while the bias increases. The RMSE decreases and the bias increases more quickly for the white oak than for the slash pine, as can be seen in the higher axis intercept of the RMSE curve.

A comparison profile for loblolly pine using the different shape parameters can be found in Fig. 2. From the intercepts of the RMSE curves we can see that the ptf fits loblolly pine better with  $c = 3$  than with  $c = 4$ , again the bias increases more quickly for the poorer fit, but the RMSE seems to decrease at about the same rate. We can see here the cost of using an inappropriate ptf.

We also examined diagnostic graphs to better understand the relative behaviors of IS and CIS. A

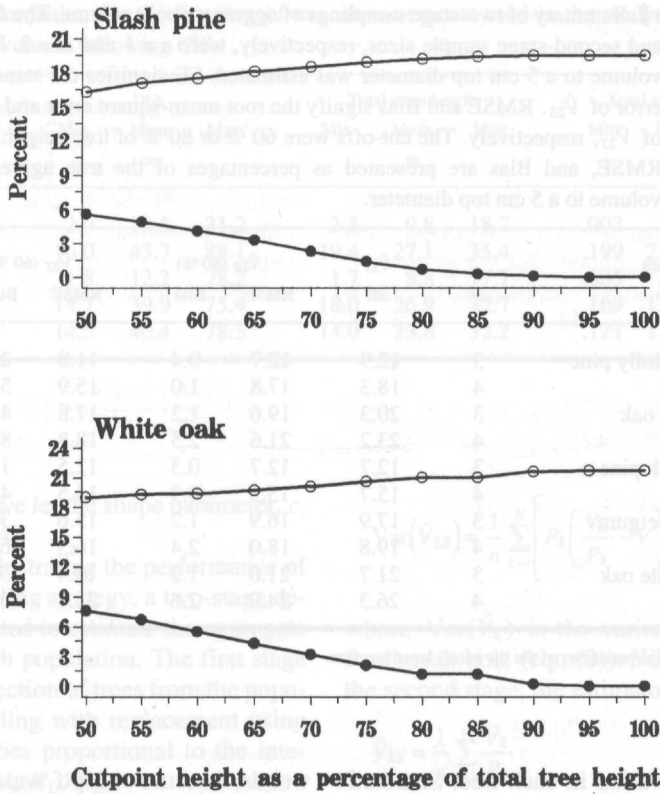


Fig. 1. Profile of the bias (bullet) and root mean-square error (circle) of  $\hat{V}_{2S}$  following CIS. Results are expressed as a percentage of aggregate bole volume to a 5 cm top diameter. The shape parameter of the proxy function was set to  $c = 3$ .

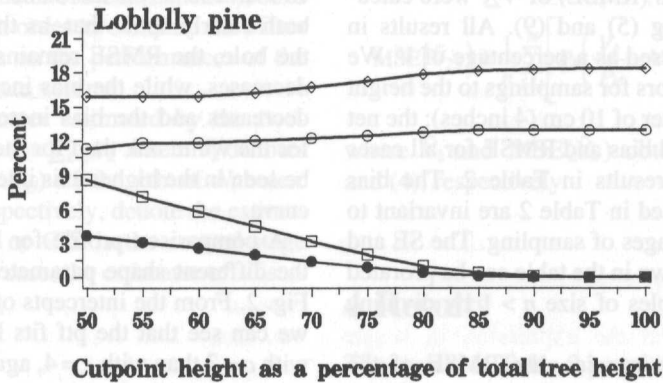
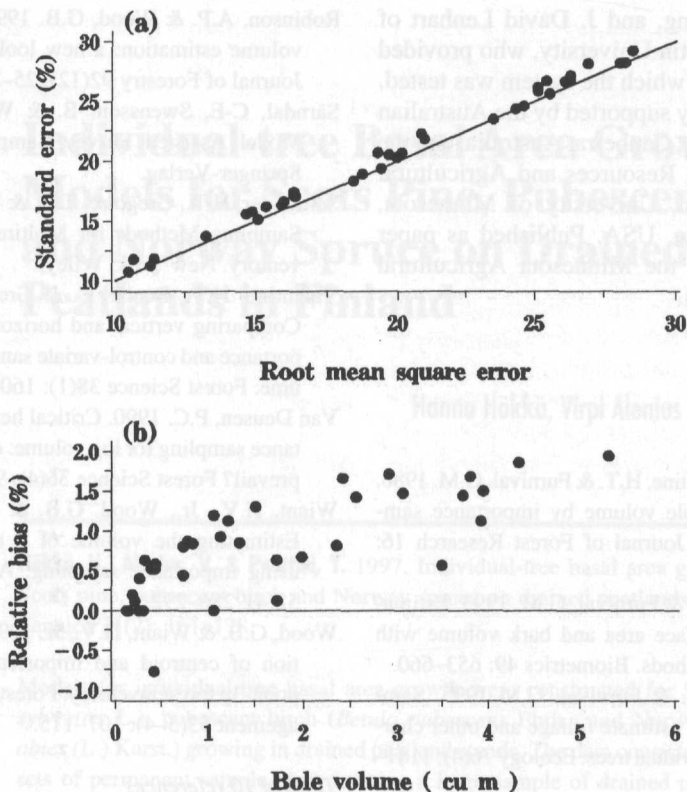


Fig. 2. Profile of the bias when  $c = 3$  (solid dot) and  $c = 4$  (square) of  $\hat{V}_{2S}$  following CIS. Also shown are the root mean square error when  $c = 3$  (circle) and  $c = 4$  (diamond). Results are expressed as a percentage of aggregate bole volume to a 5 cm top diameter.



**Fig. 3.** Estimation of white oak bole volumes. In (a), the standard error of  $\tilde{V}(H_L, H_U)$  versus the standard error of  $\hat{V}(H_L, H_U)$ . In (b), the bias of  $\tilde{V}(H_L, H_U)$  versus  $V(H_L, H_U)$ . The shape parameter of the proxy function was set to  $c = 3$ .

sample pair of graphs, for white oak, are displayed in Fig. 3. Fig. 3a shows that on an individual bole basis, the relative RMSE of  $\tilde{V}(H_L, H_U)$  is generally a bit smaller than standard error of  $\hat{V}(H_L, H_U)$ . Finally, we note with Fig. 3b that the maximum bias of  $\tilde{V}(H_L, H_U)$  is less than 2.0% and that as bole size increases, the bias increases for this particular choice of ptf.

## 5 Discussion

When the cut-off is established relatively high on the bole, the bias of total bole volume estimation by  $\tilde{V}_{2S}$  appears to be small or negligible, at least for the tree populations and two-stage sam-

pling strategies investigated here. We have demonstrated that CIS enables more accurate estimation of aggregate bole volume than unrestricted importance sampling in these circumstances. While we have not explicitly addressed the issue of measurement error, we anticipate that an added advantage of CIS results from the decreased measurement error resulting from the exclusion of the upper bole from the sampling frame.

## Acknowledgments

The authors are indebted to Geoff Wood of the Australian National University, whose original suggestions led to the development of cut-off

importance sampling, and J. David Lenhart of the Stephen F. Austin University, who provided the stem data upon which the system was tested.

Research partially supported by the Australian National University, Canberra, Australia, and the College of Natural Resources and Agricultural Experiment Station, University of Minnesota, St. Paul, Minnesota, USA. Published as paper no. 974420011 of the Minnesota Agricultural Experiment Station.

## References

- Gregoire, T.G., Valentine, H.T. & Furnival, G.M. 1986. Estimation of bole volume by importance sampling. *Canadian Journal of Forest Research* 16: 554–557.
- , Valentine, H.T. & Furnival, G.M. 1993. Estimation of bole surface area and bark volume with Monte Carlo methods. *Biometrics* 49: 653–660.
- , Valentine, H.T. & Furnival, G.M. 1995. Sampling methods to estimate foliage and other characteristics of individual trees. *Ecology* 76(4): 1181–1194.
- Robinson, A.P. & Wood, G.B. 1994. Individual tree volume estimation: a new look at new systems. *Journal of Forestry* 92(12): 25–29.
- Särndal, C-E, Swensson, B. & Wretman, J. 1992. *Model Assisted Survey Sampling*. New York, Springer-Verlag.
- Schreuder, H.T., Gregoire, T.G. & Wood, G.B. 1993. *Sampling Methods for Multiresource Forest Inventory*. New York, Wiley.
- Valentine, H.T., Bealle, C. & Gregoire, T.G. 1992. Comparing vertical and horizontal modes of importance and control-variate sampling for bole volume. *Forest Science* 38(1): 160–172.
- Van Deusen, P.C. 1990. Critical height versus importance sampling for log volume: does critical height prevail? *Forest Science* 36(4): 930–938.
- Wiant, H.V., Jr., Wood, G.B. & Miles, J.A. 1989. Estimating the volume of a radiata pine stand using importance sampling. *Australian Forestry* 52(4): 286–292.
- Wood, G.B. & Wiant, H.V., Jr. 1992. Test of application of centroid and importance sampling in a point-3p forest inventory. *Forest Ecology and Management* 53(3–4): 107–115.

*Total of 10 references*