

LARGE-SCALE ESTIMATION OF FOREST ATTRIBUTES FOR SCATTERED TREES

F. Baffetta¹, L. Fattorini¹, P. Corona²

¹*Dipartimento di Metodi Quantitativi, Università di Siena, P.za S. Francesco 8, 53100 Siena, Italy,* ²*Dipartimento di Scienze dell'Ambiente Forestale e delle sue Risorse (DISAFRI), Università della Tuscia, via San Camillo de Lellis, 01100 Viterbo, Italy*

Let \mathbf{U} be a population of trees scattered outside forests on a delineated study area \mathcal{A} .

Denote by y_j the value of a forest attribute such as timber volume or basal area for the j -th tree and suppose that the population abundance, say A and the population total

$$T = \sum_{j \in \mathbf{U}} y_j$$

be the interest quantities to be estimated. Fattorini et al (2006) propose a three-phase sampling strategy for estimating forest attributes. The estimator here proposed and their properties are completely quoted from that paper.

In the first phase, in order to perform the tessellation stratified sampling, the study area is covered by a region, say $\mathcal{R} \supset \mathcal{A}$ of size R and constituted by N non-overlapping quadrats Q_1, \dots, Q_N of equal size and such that $Q_i \cap \mathcal{A} \neq \emptyset$ for all $i = 1, 2, \dots, N$. Then a point, say \mathbf{p}_i , is randomly selected within each quadrat $i = 1, 2, \dots, N$, in such a way that a discrete population of n points, say $\mathbf{P} = \{\mathbf{p}_1, \dots, \mathbf{p}_N\}$ is obtained. If each sample point were visited on the ground and a plot of size a were constructed around each point and all the trees lying within the plot was recorded and measured, then for each point i a sample of trees, say \mathbf{U}_i , were obtained. Accordingly, if for each point $i = 1, 2, \dots, N$ the quantity

$$\hat{T}_i = \frac{R}{a} \sum_{j \in \mathbf{U}_i} y_j$$

was computed, their mean

$$\bar{T}_1 = \frac{1}{N} \sum_{i=1}^N \hat{T}_i \quad (1)$$

constituted the one-phase unbiased estimator of T . Moreover, a conservative estimator of the sampling variance of (1) was given by

$$V_{1T}^2 = \frac{1}{N(N-1)} \sum_{i=1}^N (\hat{T}_i - \bar{T}_1)^2 \quad (2)$$

Since all the N plots cannot be visited a second phase of sampling is necessary. In the second phase the population of first-phase point \mathbf{P} is partitioned into L strata, say $\mathbf{P}_1, \dots, \mathbf{P}_L$ of sizes N_1, \dots, N_L and a sample of points, say $\mathbf{S}_l \subset \mathbf{U}_l$, of size n_l is selected from each stratum $l = 1, \dots, L$ by means of simple random sampling without replacement (SRSWOR). Accordingly, if the plots corresponding to all the points selected in the second phase were visited, the quantity

$$\bar{T}_2 = \sum_{l=1}^L w_l \bar{T}_{2l} \quad (3)$$

where $w_l = N_l / N$ and

$$\bar{T}_{2l} = \frac{1}{n_l} \sum_{i \in \mathbf{S}_l} \hat{T}_i$$

constituted the two-phase unbiased estimator of T . Moreover, a conservative estimator of the sampling variance of (3) was given by

$$V_{2T}^2 = \frac{1}{N-1} \left\{ \sum_{l=1}^L w_l (N_l - 1) \frac{s_{2lT}^2}{n_l} + \sum_{l=1}^L w_l (\bar{T}_{2l} - \bar{T}_2)^2 \right\} \quad (4)$$

where

$$s_{2IT}^2 = \frac{1}{n_l - 1} \sum_{i \in \mathcal{S}_l} (\hat{T}_i - \bar{T}_{2l})^2$$

Note that if $y_j = 1$ for each $j \in \mathbf{U}$, then T reduces to the population abundance A . In this case the \hat{T}_i s reduces to $\hat{A}_i = Rd_i$, where d_i denotes the density of trees lying within plot i , and (3) provides a two-phase estimator of the population abundance, say \bar{A}_2 . Thus, in this case

$$\bar{A}_2 = \sum_{l=1}^L w_l \bar{A}_{2l}$$

where

$$\bar{A}_{2l} = \frac{1}{n_l} \sum_{i \in \mathcal{S}_l} \hat{A}_i$$

and

$$V_{2A}^2 = \frac{1}{N-1} \left\{ \sum_{l=1}^L w_l (N_l - 1) \frac{s_{2lA}^2}{n_l} + \sum_{l=1}^L w_l (\bar{A}_{2l} - \bar{A}_2)^2 \right\}$$

where

$$s_{2lA}^2 = \frac{1}{n_l - 1} \sum_{i \in \mathcal{S}_l} (\hat{A}_i - \bar{A}_{2l})^2$$

Usually, all the n second-phase points cannot be visited on the ground. In most situations only the number of trees lying within the second-phase plots are recorded from orthophotos, in such a way that only the two-phase estimate of abundance \bar{A}_2 is computed in the second phase. Usually, a third phase of sampling is necessary to estimate T . Thus, a sample of points, say $\mathbf{G}_l \subset \mathcal{S}_l$, of size m_l is selected from each second-phase sample \mathcal{S}_l by means of SRSWOR. Once the plots corresponding to all the points selected in the third phase are visited, the quantity

$$\bar{T}_3 = \sum_{l=1}^L w_l \bar{T}_{3l} \quad (5)$$

where

$$\bar{T}_{3l} = \frac{1}{m_l} \sum_{i \in G_l} \hat{T}_i$$

constituted the three-phase unbiased estimator of T . Moreover, a conservative estimator of the sampling variance of (5) is given by

$$V_{3T}^2 = \frac{1}{N-1} \left\{ \sum_{l=1}^L w_l (N_l - 1) \frac{s_{3lT}^2}{m_l} + \sum_{l=1}^L w_l (\bar{T}_{3l} - \bar{T}_3)^2 \right\} \quad (6)$$

where

$$s_{3lT}^2 = \frac{1}{m_l - 1} \sum_{i \in G_l} (\hat{T}_i - \bar{T}_{3l})^2$$

In order to estimate the population mean $\bar{Y} = T/A$, a very natural procedure is to estimate A by the first two phases of sampling, to estimate T by the three phases of sampling and then taking the ratio

$$\hat{\bar{Y}} = \bar{T}_3 / \bar{A}_2 \quad (7)$$

as the estimator of \bar{Y} . From the results of Fattorini et al (2006), (7) turns out to be approximately unbiased with an approximately conservative variance estimator

$$V_{\bar{Y}}^2 = \frac{V_{3T}^2}{\bar{A}_2^2} - 2 \frac{\bar{T}_3 C_{3AT}}{\bar{A}_2^3} + \frac{\bar{T}_3^2 V_{2A}^2}{\bar{A}_2^4} \quad (4)$$

where C_{3AT} denotes the three-phase estimator for the covariance between \bar{A}_2 and \bar{T}_3 , which is given by

$$C_{3AT} = \frac{1}{N-1} \left\{ \sum_{l=1}^L w_l \frac{N_l - 1}{n_l - 1} \frac{1}{m_l} \sum_{i \in G_l} (\hat{A}_i - \bar{A}_{2l}) \hat{T}_i + \sum_{l=1}^L w_l (\bar{A}_{2l} - \bar{A}_2) (\bar{T}_{3l} - \bar{T}_3) \right\}$$

References

Fattorini L, Marcheselli M, Pisani C (2006) A three-phase sampling strategy for large-scale multiresources forest inventories. *Journal of Agricultural, Biological and Environmental Statistics*, 11, 296-216.