Stratified sampling compared with two-phase stratified cluster sampling for timber volume estimation

D. A. Stellingwerf[1] and Sann Lwin[2]

[1] Forest Department, ITC, P.O. Box 6, 7500 AA Enschede, Netherlands
[2] Forest Department, Burma

Summary

On a rectangular area of 12 855 ha, comprising mainly Norway spruce in pure and mixed stands with some non-forest areas, two methods for determining standing timber volume were tested. Only spruce older than 40 years was included in the tests. Orthophotos were used, scale 1:10 000.

A stratified cluster sampling technique was used for method 1, estimating area proportions of spruce in each of 72 blocks (ca. 178.5 ha each). Eight sub-blocks were taken per block and the presence or absence of spruce was determined in two randomly selected clusters of five circular 0.05 ha plots per sub-block on the ortho-photos.

The geometric regression equation $y = 0.9511\ x^{0.83627}$ (x = crown cover percentage from 72 photo plots and y = volume in m³/plot from 72 field plots) was determined.

Total volume and the number of man-days for a standard error of 3 % were calculated from 720 photo plots.

Method 2 used area delineation and measurement on the orthophotos and a proportional stratified random sampling method for the determination of the spruce volume. For a standard error of 3 %, 55 plots were involved.

The first method took 23 % less office time but this advantage was lost in subsequent fieldwork. With only 360 photo plots and 36 field plots, 39 % less office time was used, which was reduced to 14 % in subsequent fieldwork. Method 2, with sampling errors only in volume, is preferred over method 1 which has sampling errors in both area and volume.

Introduction

The volume of standing timber in a forest is usually determined by means of a sampling technique, and there are several such techniques in use. They usually refer to the collection of data for calculating the volume of the trees per unit area, but they

can also be used to determine the area in which the trees grow.

When aerial photographs are used, proper film and scale are required to distinguish the individual specimens of one or more species. The usual procedure for area determination on aerial photographs is simply the delineation and measurement of the area in which the species occurs. Enumeration in field sample plots, the location of which are randomly selected in these delineated areas, will lead to the volume of the timber of the species and to its standard deviation and the standard error of the mean volume.

A very useful sampling technique is stratification of the forest area in which the species occurs. Correct stratification will result in a lower standard deviation of the volume than when no stratification is used. This is important because the lower standard deviation results in a lower number of sample plots to be enumerated to achieve a required standard error of the mean timber volume. This means a saving in man days. The delineation of the strata areas must be done correctly on the aerial photographs so that the lowest possible variance of the timber volume results. Moreover, the boundaries of the strata must be transferred from aerial photographs to maps to adjust incorrect areas which otherwise result from the perspective projection of the aerial photographs. When orthophotos are used, the areas to be measured are always correct.

Given the amount of time involved in this procedure, it seemed worthwhile to test whether another method of area determination – other than conventional delineation and area measurement – would reduce the number of man days involved. We therefore investigated the determination of area proportion of one species in rectangular blocks using plots in a stratified cluster sample.

With respect to determining the volume of standing timber, it was also worthwhile to test whether different sampling techniques used in combination with one of the area determination methods would result in a saving of time. The examples given below describe two of these combinations. For Method 1, the area proportion determination with a random cluster sample per sub-block is combined with the determination of the volume by means of a two-phase sampling technique, with circular 0.05-ha orthophoto plots for phase 1 and circular 0.05-ha field plots for phase 2. For Method 2, the area delineation and area measurement technique is combined with the determination of the volume of the standing timber by means of a stratified random sampling technique with only circular 0.05-ha field plot data.

It is important to note that orthophotos are used in both examples. They have a scale of 1:10 000 and were constructed from 1:30 000 scale black-and-white infrared aerial photographs. The orthophoto and the adjacent photographically enlarged (3×) original photographs were interpreted under a mirror stereoscope, using cut-up parts of the enlarged photos. As noted above, use of the orthophotos makes the transfer of boundaries from photo to map superfluous.

The study area was a rectangular area covering 12 855 ha in an intensively man-

aged forest in Austria, located in the state of Upper Austria, at a mean elevation of approximately 650 m above sea level. The forest cover consists mainly of Norway spruce (*Picea abies*). Beech (*Fagus silvatica*) also occurs in pure stands in smaller areas and in mixtures with spruce and silver fir (*Abies alba*). Non-indigenous species such as white pine (*Pinus strobus*) and European larch (*Larix decidua*) are also found in mixtures with spruce, but they occur far less frequently. Regeneration of trees occurs naturally in almost all areas.

For the purpose of this article, unless otherwise stated, spruce refers to Norway spruce more than 40 years old.

Description of the methods

Method 1. Stratified cluster sampling for area proportion estimation with orthophoto plots and two-phase stratified cluster sampling with photo and field plots for volume estimation. The separate steps include the following.

I *Office work*
(1) *Drawing the rectangular external boundaries, comprising 12 855 ha, followed by drawing 72 equally sized and shaped rectangular blocks on the orthophotos.* Each block covers 178.54 ha and is subdivided into eight equally sized and shaped rectangular sub-blocks. The division of the total area into 72 blocks and of each block into eight sub-blocks (see Fig. 1) is a rough stratification. This stratification is usually less effective with respect to the volume variance reduction than the division of the
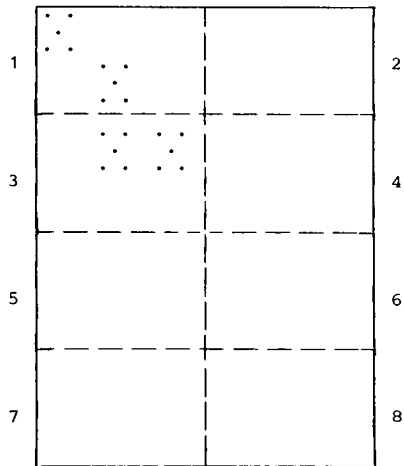

Fig. 1. One of the blocks with eight sub-blocks. Sub-blocks with two randomly located clusters of 5 plots each (only indicated in the sub-blocks 1 and 3). The centres of the circular plots, which cover 0.05 ha each, are indicated. The plots are used for the area proportion determination of spruce.

spruce stands into the two density strata used in the second method. It is far less time-consuming however (time = 0.50 man-days).

(2) *Delineation on the orthophotos of the areas occupied by young spruce stands ($\leqslant$ 40 years) with existing management maps.* These stands cover approximately 30 % of the total forest area (time = 2.50 man-days).

(3) *Determination of the area proportion of spruce forest by means of sample plots on orthophotos.*

– The positions of two clusters, each comprising five circular plots of 0.05 ha each, are randomly selected in each sub-block. In total, there are 1152 clusters (2 clusters × 8 (sub-blocks) × 72 (blocks)) and 5760 plots (1152 × 5) (see Fig. 1). The selection of the positions takes 2.30 man-days.

– The presence or absence of spruce trees in each of the 5760 plots is noted under a stereoscope (time = 2.30 man-days).

– If in general we write k = total number of plots per cluster,

$\qquad$ k1 = number of plots which contain spruce out of k plots

$\qquad$ and m = number of clusters which are randomly selected in each block,

then the total number of plots which contain spruce, can be written as follows: $\overset{m}{\Sigma}$ k1 = $\overset{8}{\Sigma}\overset{2}{\Sigma}$ k1, since there are 2 clusters per sub-block and 8 sub-blocks per block. Block 1, for example, then gives the following plot data in the cluster (the data refer to the presence (maximum 5) or absence (0) of spruce):

Block 1 total = (0+0)+(0+0)+(2+0)+(4+0)+(0+0)+(1+4)+(0+0)+(0+0)

$\qquad$ $\overset{8}{\Sigma}\overset{2}{\Sigma}$ k1 = 2+4+5 = 11.

The proportion pi of each block is then calculated from $\overset{8}{\Sigma}\overset{2}{\Sigma}$ k1/$\overset{8}{\Sigma}\overset{2}{\Sigma}$ k, in which for each block $\overset{8}{\Sigma}\overset{2}{\Sigma}$ k = $\overset{8}{\Sigma}\overset{2}{\Sigma}$ 5 = 80, since there are 5 plots per cluster, 2 clusters per sub-block and 8 sub-blocks per block. For block 1, p1 = 11/80 = 0.1375.

The variance of the proportion p can be calculated dividing the variance of the block total by the square of the total number of plots. Hence:

V(p) = V($(\overset{8}{\Sigma}\overset{2}{\Sigma}$ k1$)/(\overset{8}{\Sigma}\overset{2}{\Sigma}$ k$)^2$).

For block 1 the variance of the block total is

$V(11) = (0–0)^2 + (0–0)^2 + (2–0)^2 + (4–0)^2 + (0–0)^2 + (1–4)^2 + (0–0)^2 + (0–0)^2$

$\qquad$ = 4 + 16 + 9 = 29.

and V(p1) = V(11/80)

$\qquad$ = $V(11)/80^2$

$\qquad$ = 29/6400

$\qquad$ = 0.004531.

The same calculation procedure is used for the remaining 71 blocks. The resulting proportions are applied block by block to the data of the second activity (see step 4 below), because the variance of the stocking proportions vary greatly from block to block.

The calculations take 0.50 man-day.

(4) *Volume determination of spruce with a two-phase sampling technique.*
– Two clusters of five plots each are randomly placed in each block. These plots, which also cover 0.05 ha, serve for the indirect determination of the volume of spruce timber. Though these clusters are selected independently of the ones which are used for the area proportion determination of the spruce (see step I (3)), they could also have been selected dependently. In total there are 72(2)(5) = 720 photo plots involved. The selection of these cluster takes 1.40 man-days.
– The determination of the crown cover percentage of the spruce on these 720 photo plots requires stereoscopic photo images. Parts of individual 3 × enlarged 1:30 000 scale photos, together with the respective 1:10 000 scale orthophotos, are therefore used under a mirror stereoscope. 10 % classes are taken for the crown cover, except for the lower classes which each cover 5 %. The crown cover determination takes 3.60 man-days.
– One of the ten photo plots per block is selected. In total therefore 72 plots are selected, which serve as field plots on which the volume of the spruce trees is determined. The selection is made in such a way that on the 72 plots the entire range of crown cover percentages is present. The calculated plot volumes are used for the calculation of the exponent and the coefficient of a regression equation (see step III (1)). The selection of the 72 plots takes 0.10 man-days.

II *Fieldwork*
Some assumptions are made for the calculation of the required number of man days involved in the completion of the various steps of the field work. These assumptions are:
– The field crew consists of two men.
– One man-day is equivalent to 8 man-hours.
– The enumeration of one circular 0.05-ha plot in the field takes 0.33 crew-hours.
– Daily travel from base camp to the first plot and return is done by car at an average speed of 50 km/h. The average daily distance covered by car is 13 km (2 × 6.5). This involved 0.065 man-days, or (13/50)/(2/8).
– Travel from plot to plot is done on foot at an average speed of 1.8 km/h. A systematic distribution of the plots is considered for calculating the average distance between the plots. The average distance d in metres is calculated from d^2 = (area in m)/(number of plots). The total distance which is covered on foot is then m(d) metres, in which m is the number of plots.
  The average distance between the 72 plots is 1336.2 metres or (12 855 × 10^4/72)$^{0.5}$.
  The number of plots that can be completed per man-day is 3.488. The following equation is used for this calculation:
0.065 + m(1336.2/1800)(2/8) + m(0.33)(2/8) = 1,
in which the first term indicates the time involved in travelling by car from base

camp to first plot and return, the second term the time for walking from plot to plot, and the third term the time for the enumeration of the plots in the field. For the completion of m = 72 plots 20.64 man-days (= 72/3.488) are needed.

III *Final calculations*

(1) The geometric regression equation indicating the relation between field volume (y) (Method 1, step II) and photo crown cover percentage (x) (Method 1, step I (4)) is found to fit the 72 plot data in the best way. In general such a geometric equation can be written $y = b x^c$. After calculation of c and b, the equation reads $y = 0.9511$ $x^{0.82627}$. The variance of the regression coefficient b is calculated from

$$V \; b = [s_{yx^c}]^2 / \overset{72}{\underset{}{\Sigma}} \, x^c$$
$$= 1.76611/1568.72209 = 0.001126.$$

The completion of this calculation takes 0.50 man-day.

(2) The weighted means of the x^c values per block and their variances are calculated from 2×5 plots per block. The data of block 1 may serve as an example. The individual x^c values of 5 plots in cluster 1 are 0, 0, 0, 0 and 0 respectively. The cluster total which is called 1a is 0. In cluster 2 of block 1, the x^c values of the 5 plots are 0, 35.425, 31.474, 0 and 18.872 respectively. The cluster total 1b = 85.771.

The weighted mean of the x^c plot values for each block is from $\overline{x^c}$ = t/(n1+n2), in which n1 = number of stocked plots in 1, n2 = number of stocked plots in cluster 2 and t = 1a+1b. For block 1, t = 85.771, n1 = 0 and n2 = 3. Hence:

$\overline{x^c}$ = 85.771/(0+3) = 28.590.

Substituting x^c of the regression equation by $\overline{x^c}$ of the weighted mean per block results in the mean volume per block. For block 1 $\overline{x1^c}$ = 28.590, and therefore after substitution:

$\overline{y1}$ = 0.9511 $\overline{x1^c}$

= 0.9511 (28.590) = 27.192 m^3/0.05 ha.

The variance of the weighted mean for each block, $(\overline{x^c})$, when the two selected clusters have stocked plots, is calculated according to

$V(\overline{x^c})$ = [(ia)^2/n1+(ib)^2/n2−(ia+ib)^2/(n1+n2)/(n1+n2),

in which i is the block number.

The variance of the weighted mean for a block, in which one of the clusters has no stocked plots, is written as follows:

$V(\overline{x^c})$ = $[\overset{n}{\underset{}{\Sigma}}(x^c)^2 - (\overset{n}{\underset{}{\Sigma}} x^c)^2/n]/(n(n-1))$.

For block 1, 1a = 0, 1b = 85.771, n1 = 0 and n2 = 3. Hence the second variance formula gives:

$V(\overline{x^c})$ = $[\overset{3}{\underset{}{\Sigma}}(x^c)^2 - (\overset{3}{\underset{}{\Sigma}} x^c)^2/3]/(3(3-1))$

= [2601.695−(85.771)^2/3]/6 = 24.91221.

The variance of the block mean can be calculated by using the propagation of er-

rors: $V \bar{y} = (\overline{x^c})^2 V b + b^2 (V (\overline{x^c}))$. For block 1, $V\overline{y1} = 28.590^2(0.001126) + 0.9511^2(24.91221) = 23.45575$.

The mean volumes and their variances of the other 71 individual blocks are calculated with the same procedure. These values must be adjusted to the average plot over both stocked and unstocked plots with spruce, since these values refer to the weighted means and their variances. The adjustment is accomplished by multiplying the weighted mean per plot of a block $(\overline{yi})$ with the respective pi value. For block 1, p1 = 0.1375 and $\overline{y1}$ = 27.192. Thus:

$\overline{y1}$p1 = 27.192×0.1375 = 3.739 m^3/0.05 ha.

The variances of the adjusted mean volumes result when the propagation of errors is applied:

$V(\overline{yi}\, pi) = pi^2\, V\overline{yi} + (\overline{yi})^2 V\, pi$.

For block 1, p1 = 0.1375, Vp1 = 0.004531, $\overline{y1}$ = 27.192 and $V\overline{y1}$ = 23.45575. Hence:

$V(\overline{y1}p1) = V(3.739) = 0.1375^2(23.45575) + 27.192^2(0.004531) = 3.79370$.

The sum of all 72 blocks is $\sum^{72} (pi\overline{yi}) = 811.222$ m^3/0.05 ha and its variance $\sum^{72} V(pi\overline{yi}) = 564.00141$. The standard error $S(\overline{yi}pi) = (564.00141)^{0.5} = 23.7490$ m^3/0.05 ha.

After completion of the previous calculations the volume and its standard error for the total 12 855 ha can be determined.

Total volume: yt = 12 855×811.222/(72×0.05) = 2 896 739 m^3.

Standard error: Syt = 12 855 (23.7490)/(72(0.05)) = 84 803 m^3.

In order to make a comparison with Method 2, this standard error must be expressed in m^3 per hectare spruce forest, because s1 and s2 of Method 2 are expressed in this unit. Therefore, the total area covered by spruce in pure and mixed stands is calculated as the sum of the products of the block areas and the respective pi values. This sum is 6 867.0637 ha in which the block area of 178.54 ha is constant. The standard error is therefore 84 803/6 867.0637 = 12.3492 m^3/ha. This value is also used in Method 2, step I(6).

The calculations take 1.50 man-days.

The total number of man-days for Method 1 for the categories office, field and final calculations are 13.10, 20.64 and 2.00 respectively. In total *35.74 man-days* are needed.

Method 2. Volume determination using a stratified random sampling technique with field plot data; area determination by delineation on orthophotos. The separate steps involved in this method include the following.

I *Office work*
(1) Drawing the rectangular boundary of the total area, covering 12 855 ha (time = 0.14 man-days).

(2) Delineation of the exterior forest boundary plus delineation of the young spruce stands ($\leqslant$ 40 years), making use of the existing management maps on which the boundary of the latter category is indicated (time = 3.0 man-days).

(3) The areas with pure broad-leaved species, covering 3 % of the total forest area, and the non-forest areas within the forest, covering 10 % of the area, are delineated (time = 1.00 man-days).

(4) Two strata, density 1 and density 2 of the pure and mixed spruce stands, are distinguished and their areas are delineated (time = 4.50 man-days).

(5) Area determination by planimeter of the various areas, including stratum areas 1 and 2, followed by calculation of the area proportions p1 (= 0.49) and p2 (= 0.51) of strata 1 and 2, respectively (time = 8.4 man-days).

(6) A limited number of 10 circular 0.05-ha plots per stratum are enumerated in the field to calculate the standard deviations s1 and s2 for stratum 1 and stratum 2, respectively. The standard deviations are s1 = 83.9800 and s2 = 98.5000.

A comparison of Methods 1 and 2 is possible if the standard errors are made equal. The standard error calculated from method 1 is used to calculate the required number of plots for Method 2. The calculated standard error (S) of Method 1 is 12.3492 m^3/ha. The number of plots (m) in Method 2, which will lead to the same standard error, is calculated from m = $(\overset{2}{\Sigma}\text{pisi})^2/S^2$, because a proportional distribution of the plots over the two strata will be made.

The number of plots follow from m = $(0.49(83.9800)^2$ + $0.51(98.5000)^2)/(12.3492)^2$ = 55. The number of plots for stratum 1 is m1 = 55(0.49) = 27, and for stratum 2, m2 = 28. This means that for the strata 1 and 2 the additional plots which must be located randomly and enumerated in the field are 27 − 10 = 17 and 28 − 10 = 18 respectively.

The location of these plots on the orthophotos takes 0.06 man-day.

II *Fieldwork*

The same assumptions which are made for Method 1 are also made for Method 2. The average distance between the 55 field plots is 1528.8 m, or $(12\ 855 \times 10^4/55)^{0.5}$.

The number of plots that can be finished per man-day is 3.1713. This is calculated with the equation:

$0.065 + m(1\ 528.81/1\ 800)(2/8) + m(0.33)(2/8) = 1$.

For the completion of 55 plots 55/3.1713 = 17.34 man-days are needed.

III *Final calculations*

Final calculations are needed for total volume calculation and for the activities as described in step I(6). Time involved 1.50 man-days.

The total number of man-days for Method 2 for the office, field and final calculations are 17.10, 17.34 and 1.50 man-days, respectively. In total, *35.94 man-days* are needed.

Comparison of the two methods

For both methods, the 1:10 000 scale orthophotos (made from 1:30 000 scale aerial photographs) proved to be very useful because the individual spruce trees were detectable in the stereo images. The 1:10 000 scale for the orthophotos was selected because the management maps have the same scale. The original 1:30 000 photo scale is in practise the smallest scale that can be used. It makes the production of the 1:10 000 orthophotos less costly than when starting from larger scale aerial photographs and still guarantees sufficient detail for management purposes to be visible in the final orthophotos.

The comparison of the number of man-days involved in the two methods, taking the same standard error (= 12.3492 m³/ha), gives interesting results. These data are given in Table 1.

The result shows that under the given circumstances both methods consume practically the same number of man-days. Method 1 requires fewer office man-days as expected. These savings, however, are lost by the extra man-days spent in the field. (This is obvious because the number of field plots used for the regression equation for Method 1 is 72 against 55 field plots for Method 2.)

The extra sampling error in area of Method 1 is a detrimental factor which diminishes the advantage which a two-phase sampling method usually has over a stratified sampling method. Also the stratification in blocks which is used for Method 1 is usually less effective with respect to the reduction of the variance than the stratification used for Method 2. When the number of blocks for Method 1 is reduced from 72 to 36, with 16 sub-blocks per block and 2 clusters of 5 plots each per sub-block are taken, then the number of plots for the determination of the area proportion is still

Table 1. Number of man-days for office, fieldwork and additional calculations and number of field and photo plots for Methods 1 and 2.

Method	Number of man-days				Number of plots	
	office	field	final calculations	total	field	photo
1	13.10	20.64	2.00	35.74	72	720
2	17.10	17.34	1.50	35.94	55	0

Table 2. Number of man-days for office, fieldwork and final calculations and number of field and photo plots for modified Methods 1 and 2.

Method	Number of man-days				Number of plots	
	office	field	final calculations	total	field	photo
1	10.40	13.28	1.75	25.43	36	360
2	17.07	11.13	1.50	29.70	27	0

5760. The number of photo plots for determination of the crown cover percentage is then 360 instead of 720. Taking for the regression equation calculation 36 plot data instead of 72, then the resulting standard error of the total volume of the total area is Syt = 120 746.3 m^3. Expressing this value in m^3 per hectare spruce forest, S$\bar{y}$ = 120 746.3/6 867.0637 = 17.5834 m^3/ha.

When this same standard error is taken for Method 2, a total of 27 0.05-ha field plots are needed. The number of man-days for the two methods is calculated in the same way to that described above. The results are shown in Table 2.

Again, fewer office man-days are shown for Method 1. The gain, as compared to the previous conditions, is increased because now only 360 instead of 720 photo plots for crown cover percentage determination are involved. This gain of 39 % is now only partly neutralized by the extra field man-days which are needed to enumerate the 36 regression plots, and the final gain is still approximately 15 %.

An advantage of Method 1 over Method 2, is that no extra delineation is to be made except the linear boundaries of the blocks and the sub-blocks. The determination of the presence or absence of spruce in the plots on the photographs under a stereoscope is done very quickly (5760 plots are completed in 2.3 man-days). Another advantage is the lower total number of man-days needed when the number of regression plots in the field and the number of photo plots are reduced to 36 and 360, respectively. A disadvantage of Method 1 is that this method does not allow the calculation of the number of clusters that will be needed for the area and the volume determination for a pre-defined standard error. This is caused by the variance of the sum of the products per block of the area proportion and the respective mean volume ($\sum^{2}$ (Vp$\bar{i}$y$\bar{i}$)). For Method 2, this number is much easier to calculate, because the areas of the strata are free of sampling error. It is for this reason that Method 2 is preferred over Method 1, notwithstanding the advantage of the simpler area proportion determination of Method 1 and the gain in man-days for 36 blocks, 36 regression plots in the field and 360 photo plots.

Bibliography

Sann Lwin, 1984. The application of two-stage sampling with probability proportional to size and two-phase stratified cluster sampling using aerial photography and Landsat imagery in forest inventory. ITC thesis. ITC, Enschede, Netherlands, 238 pp.

Schumacher, F. X. &R. A. Chapman, 1954. Sampling methods in forestry and range management. Duke University, School of Forestry, Bulletin 7 (revised), 221 pp.

Trân-Châu, L., 1977. Efficiencies of several sampling methods. ITC thesis. ITC, Enschede, Netherlands, 75 pp.