

**SYSTEMATIC SAMPLING FOR ESTIMATING HARVEST-INDUCED
CHANGES OF A FOREST STAND**

Dennis P. Peque

College of Forestry and Natural Resources

Visayas State University

Visca, Baybay City, Leyte

ABSTRACT

This study was conducted in Compartment 2012a in Bösinghausen Forest District in Germany covering an area of 5 hectares. Twenty two sampling plots were laid out in the field following systematic sampling design.

Results showed that all estimates for all variables (e.g. tree heights, DBH, stem density, basal area and volume) under trees that were marked for cutting have higher relative standard error. This was due to higher dispersion of individual estimates in each plot. On the other hand, the simulation study shows that sampling efficiency can be achieved by increasing the sample size. When more samples are included, the relative standard error becomes low.

From this study, it can be concluded that the variability of the estimates were affected by sample size and the variability of individual units in the population or the individual estimates (in this case, estimates in each plot).

Key Words: Systematic sampling design, Simple random sampling (SRS), DBH, sampling efficiency.

INTRODUCTION

From a management point of view, a clear understanding about the status and/or conditions of the resource is necessary if appropriate actions and interventions are to be proposed or implemented. When a forest stand is adequately stocked and that the harvestable volume are sufficiently enough, then one can proposed for harvesting of the said stand. In most of the cases however, changes that have taken place in the characteristics of a forest population are often of as much interest as their present status. Estimating such changes usually requires sampling at the beginning and at the end of the period. Changes in the stand in terms of density of trees as well as stand structure are normally measured by the difference of the estimates of the present status and the estimates of the harvesting intervention. This information becomes a basis on how the stand shall be managed in the future.

The situation in the field such as the size of inventory area, spatial distribution of trees, the structure of forest and stands, as well as the inventory objectives, has led to the development of different sampling methods. It would have been convenient if there is one sampling method that is so efficient that it could be used regardless of the forest type and inventory objective (Loetsch and Haller, 1973). Unfortunately, there is none in forestry. In this study, systematic sampling was used to estimate changes in the forest stand as a result of planned harvesting intervention.

The general objective of this study was to estimate changes of the forest stand in terms of density and structure as a result of planned harvesting intervention. Specifically, it aims:

- a. To estimate the total number of stems in the study area, their diameter distribution, basal area, and stand volume;

- b. To assess changes of the above-mentioned variables of interest when trees marked for cutting are finally harvested;
- c. To compare various estimates using different sample size.

The Sampling Design

Systematic sampling

Systematic sampling is the most common sampling method used in forest inventory. The units included in a systematic sample are not selected at random but according to a pre-specified pattern. Usually the only element of randomization is in the selection of the starting point of the pattern. The selection of the first sampling unit automatically determines the position of all other units. The most common pattern is a grid having the sample units in equally spaced rows with a constant distance between units within rows (Freese, 1962). Other types of layouts are continuous strip layout and the line plot lay-out (Loetsch and Haller, 1973).

Statistical problems of systematic sampling

If we consider the systematically selected sampling points (units) as mutually independent and as located by chance and if we consequently adopt the formulas appropriate for calculating the estimate of the variance and the standard error in an unrestricted random layout of sampling units, an over-estimate of the variance and consequently of the standard error of spatially heterogeneous populations will be obtained (Loetsch and Haller, 1973). Experience suggests that in forest inventory the results of systematic samplings are more precise than those of random samplings (SRS) of equal sampling fraction. Unfortunately it is not possible in a statistically acceptable manner to ascertain the degree to which the

result is more precise. The standard error, calculated by the general formula for SRS, is no more a measure of mean square deviation of many sample means from the mean of the population, but rather a measure of a maximum deviation of which the probability of discrepancy is unknown (Loetsch and Haller, 1973).

Scheaffer et al., 1990 as cited by Shivers and Borders (1996) show that SRS estimators of variance serve as valid estimates of the variance of systematic sampling estimators for all types of populations except for periodic populations. In a periodic population the SRS variance estimators will underestimate the true variance population (Shivers and Border, 1996). Since this type of population occurs in practice, we must have to be aware of them.

The estimators

In timber inventory the population parameters of greatest interest are usually the mean per sampling unit (e.g. mean stems per hectare) and the population total. The sample estimators of these parameters and their associated variance estimators are as follows:

Sample Mean

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

Sample Variance

$$S_y^2 = \frac{1}{n-1} \left[\sum_{i=1}^n y_i^2 - \frac{\left(\sum_{i=1}^n y_i \right)^2}{n} \right]$$

where: y_i stands for the observed values or elements.

METHODOLOGY

Study site

The study area was located in Compartment 3012a within the forest area of Bösinghausen, Göttingen, Germany covering a forest stand area of 5 hectares. The forest area is characterized to be relatively flat and was intensively managed, the most dominant species of which was European Beech.

Göttingen, which is located in the state of Niedersachsen, has an average rainfall of 641 mm. The average warmest temperature in the area is 18 °C which falls in the month of July and the coldest temperature is -2 °C which usually happens in January.

Sampling design

This study employed systematic sampling design (Fig. 1). In this method, the individual sampling units were selected according to a rigid, predetermined scheme. The principal aim was to cover the population in all its parts as uniformly as possible and to achieve simplicity of layout. A total of 22 500m² plots were established located at equal distances of 50 x 50 meters within Compartment 3012a.

Sample size

Dividing the total area of Compartment 3012 which was 5 hectares by 20 yielded 2500 m², which is essentially the total area represented by each point. Taking the square root of this value would result to 50 m. This means that the sides of the square grid are measuring 50 meters. Because of the shape of the compartment and the orientation of the grid when the starting random point was located, a total of 22 plots were established in the study area.

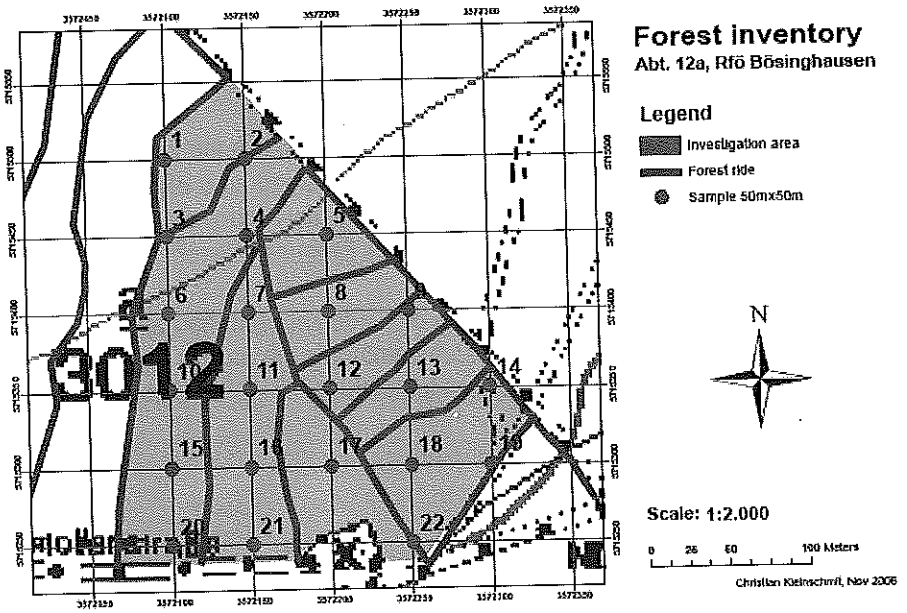


Figure 1. Study area and the systematically arranged plots

Data gathering

Following the inventory protocol, each crew was responsible for attaining the objectives of the inventory. Although, the coordination team was also required to perform the control measurements or was tasked to check the field crews in order to ensure that the inventory protocol was being carried out correctly so that the data gathered will all be reliable.

Finding the exact location of the sampling points in the field are one of the factors that influences progress of inventory or surveying work especially in

canopy-covered areas. Since at that time the Global Positioning System (GPS) device did not give a good signal, the crews decided to locate the sampling points with the aid of conspicuous landmarks such as corners of roads and/or skidding lines. After locating the sampling points, data gathering then started. All trees ≥ 7 cm diameter-at-breast height (DBH) that was within the 12.62 m radius was considered. The variables collected were species name, DBH, top heights of selected trees as well as height to tree crown base, distance of trees to plot centers and azimuths. DBH were measured by diameter tape or tree caliper. On the other hand, heights of trees were measured by Spiegel relascope, Blume-Leiss or Vortex. Tally trees or trees accounted in the area were recorded as either 0 or 1. One (1) indicates that trees are to be cut soon. Trees marked for cutting were indicated by red ribbon that was tied on their stems.

Analysis of the data

Since only three trees were measured for total height (TH) and height to the crown base (HCB) in every plot, the corresponding heights of other trees in the stand were derived using simple linear regression functions of Statistica 6.0 software. From the model, the coefficient of correlation for total height was 0.4377 and the r^2 value for the (HCB) was 0.2898. One problem with this linear model is that even a DBH = 0 would yield a height of about 14 meters which indicates that a linear height model is not appropriate. Though this problem was noticed, the model was still used considering that other height models are too complex to build and are difficult to derive. Anyway, the concern of this paper is more on the statistical issues of the sampling design rather than the height values themselves. Wood volume was calculated using HCB values and a form factor of 0.7 which are recommended for tree plantations.

An estimate of stand characteristics such as stem density, height, DBH, basal area and volume was computed using the SRS estimators. Estimates using different plot sizes were derived through simulation to compare their result from using empirical data. The result gives inference about the relationship between sample size and the efficiency of the estimate.

RESULTS

Species Distribution

The dominant species in the study area was Beech (Buche) which constitutes 89.39% of the total number of species in the stand. It is followed by Sycamore maple (4.17 %) as shown in Table 1.

Table 1. Frequency distribution of species in the stand

Species name	Proportion from the mean (trees/plot)	Proportion from the mean (trees/ha)	Proportion from the Total	Percent
Beech (Bu)	10.73	214.55	1073	89.39
Sycamore maple (Bah))	0.50	10.00	50	4.17
Ash (Es)	0.23	4.55	23	1.89
Norway maple (SAh)	0.41	8.18	41	3.41
Elm (Ul)	0.05	0.91	5	0.38
Service tree (Elsb)	0.05	0.91	5	0.38
Cherry (Kir)	0.05	0.91	5	0.38
Total	12.00	240.00	1200	100

The resulting species distribution when trees that were marked for cutting are to be finally removed is shown in Figure 2. It was observed that Beech species were mostly large enough and were ready for harvesting compared to other tree species in the stand. Since Beech was the dominant species in the area, naturally those that will be left after cutting would also be dominantly composed of Beech.

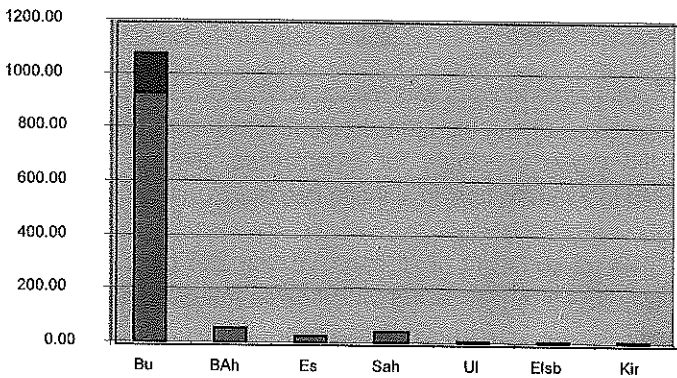


Figure 2. Proportion of species “to be cut” composed of purely Beech species

Stem density

The average number of trees in the study area was 240 trees/ha (Standard Error, SE % = 5.46%) and a total of 1200 trees for the whole stand. The average number of trees to be cut was 30 trees/ha (SE % = 15.66 %) while the remaining trees were 210 trees /ha (SE % = 6.59%). Table 2 shows the estimated change in the total number of trees in the stand.

Table 2. Present and future stem density of the stand

Variables	Present status	Planned harvest	Change
No. of trees/ha	240	30	210
Total no. of trees	1200	150	1050

Height Distribution

Total height of trees was measured from the base up to the tip of the tree. Error can occur during the measurement considering that it is difficult to find the tip of the crown especially for broadleaved species. Thus, the crews were instructed to exercise care in sighting tree tops.

The mean height of all trees, i.e. the 264 trees encountered during the study was 25.75 m (SE % = 1.24 %) while the average HCB of the trees was 12.69 m (SE % = 1.57 %). On a plot basis, the mean height was 26.02 (SE % = 7.52 %) while HCB was 12.86 (SE % = 9.38 %). While the first approach produces a very low relative error the result is not valid because the trees in the stand were considered or used as sampling/observation unit rather than using the plots themselves. This means that we changed the sampling size from 22 to 264 which, in statistical sense is not correct and therefore biased.

Again on a plot basis, the average total height and height to crown base of trees that were marked for cutting were 24.50 m (SE % = 59.76 %) and 12.49 m (SE % = 50.45 %), respectively. Table 3 shows the mean total height and height to crown base of the remaining trees in the stand. From the table it is clear that both TH and HCB of the trees to be cut have higher % SE. The reason behind this is that some plots did not contain trees marked for cutting. In statistical parlance,

this means that there is a great dispersion of the individual values (in this case individual estimates of each plot) which produces greater standard deviation or the standard error.

The data also indicate that majority of the trees in the stand belonged to height classes 25-30 m (41.29%) while 18.56% belonged to 30-35 m height classes. A good percentage (23.48%) of trees in the stand belonged to 15-20 m category.

Table 3. Average height characteristics of the trees in the study area

Status	Total Height (m)		Height to Crown Base (m)	
	mean	% SE	mean	% SE
Original	26.02	7.52	12.86	9.38
Cut	24.50	59.76	12.49	50.45
Remaining	7.88	12.58	12.58	9.88

Diameter Distribution

Diameter distribution is an indicator of the structure of the growing stock and therefore become a basis for economic decisions as well as in planning. As shown in Figure 4, majority of the trees in the study area belongs to the DBH classes 40-50 cm (37.12 %) which is followed by 10-20 cm classes (25.76%). The mean diameter of trees in the stand is 34.26 cm (SE % = 2.77 %) with 65.57 cm as the biggest diameter. The estimated mean diameter of trees that were marked for cutting is 45.29 cm (SE % = 4 %) while the mean diameter of the remaining trees in the stand is 32.69 cm (SE % = 3.09 % SE). Table 4 shows the estimates of diameters of trees in the stand on a plot basis. It can be seen that the relative standard error for DBHcut is 52.16 % which is essentially very high. The reason for this is that some plots did not have trees that were marked for cutting which translates to sampling plots having no DBH values. This resulted to greater variability of the observed values which consequently affected the relative standard error.

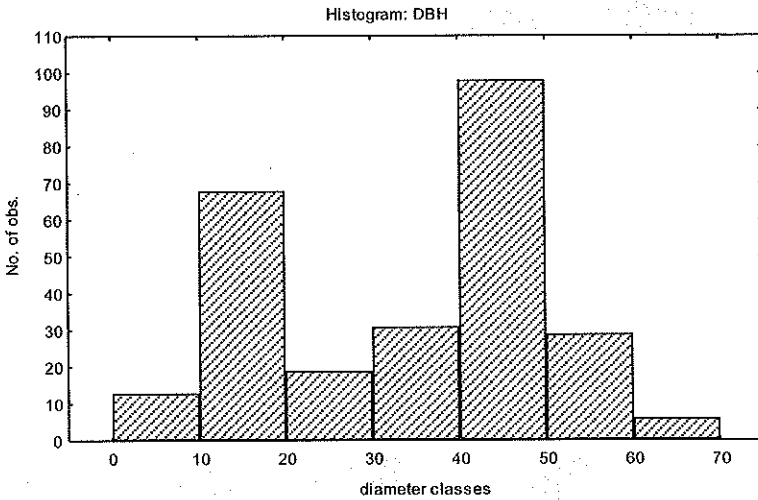


Figure 3. Diameter distribution of trees in the stand

Table 4. Mean estimates of different DBH categories

Variables	Mean (m)	Variance	Std. Error	%SE
DBHactual	35.24	33.27	5.77	16.37
DBHcut	37.91	390.88	19.77	52.16
DBHremain	33.95	34.38	5.86	17.27

Basal area and Volume

The estimated mean basal area of the trees in the stand was 22.99 (% SE = 27.58 %) m²/ha with a total of 114.95 m². Conversely, the estimated mean basal area for trees that were marked for cutting and those trees that are to remain were 4.38 m² (SE % = 61.64 %) and 27.83 m² (SE % = 26.80 %) per ha, respectively. The estimated total commercial volume of wood in the stand was 1461.85 m³ with an average of 292.37 m³/ha (SE % = 32.3 % SE) while for the remaining trees was 236.26 m³/ha (Table 5).

Table 5. The estimated commercial volume of trees in the stand.

Commercial Volume	Mean (m ³ /ha)	Total (m ³ /ha)	Std. Error	% SE
Actual	292.37	1461.85	94.44	32.30
Planned (to be cut)	49.04	245.20	36.63	74.69
Remaining	236.26	1181.3	82.30	34.84

Estimates of stem density using different sample sizes

Using the same data obtained from the stand, an estimate of stem density was determined using different sample sizes ($n= 10; 15; 20; 25; 30; \text{ and } 35$) through simulation. It was done by generating 100 random combinations of numbers from 1 to 22 which correspond to the total number of plots used in this study. If for example, 21 is generated, the actual stem density of plot 21 is being used. This procedure was being repeated 10, 15, 20, 25, 30, and 35 times. The results indicated that the mean values of the tree stems per hectare were close to the mean values computed from the empirical data. Table 6 shows that sample size of 20 (239.86 trees/ha) has almost the same estimated mean with $n=22$ (240 trees/ha) of this study. Also, from the graph that follows (Figure 4), it can be observed that the relative standard error (RSE) decreases with an increasing sample size.

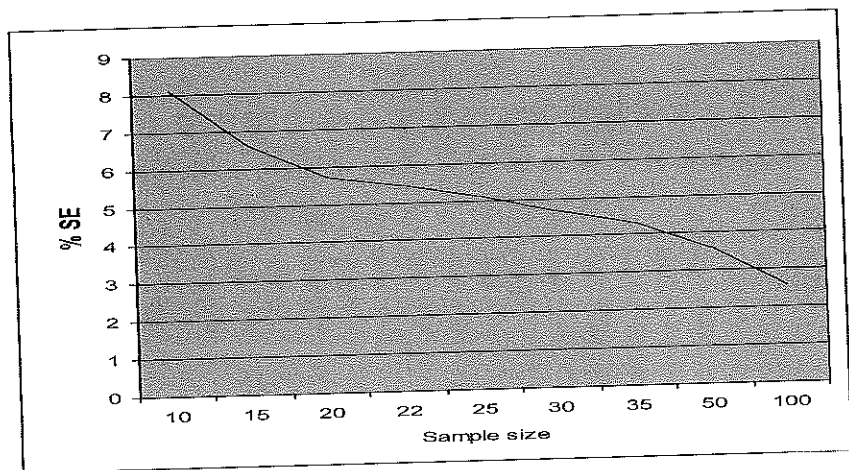


Figure 4. Graph showing the relationship of relative standard error and sample size

The relative standard error (SE %) of the number of stems per hectare was also computed using the formula as follows:

$$s_{\bar{y}} = \frac{S}{\sqrt{n}}$$

$$SE(\%) = \frac{s_{\bar{y}}}{\bar{y}} \times 100$$

where: S is the sample standard deviation (61.41), $s_{\bar{y}}$ is the standard error of the mean, and n is the desired sample size. Although the %SEs ($n_{10}=8.09\%$; $n_{15}=6.61\%$; $n_{20}=5.72\%$; $n_{25}=5.12\%$; $n_{30}=4.67\%$; $n_{35}=4.33\%$) were a little much higher than the % SE as shown in Table 6, the same trend can still be observed.

These results indicate that sampling efficiency can be achieved by increasing the sample size. The more samples are included the lower the relative standard error becomes. This goes to show that when n approaches N , $s_{\bar{y}}$ becomes zero (Kleinn, 2005).

Table 6. Estimates of the number of stems per hectare using different sample size

Sample size	Mean	Lower bound	Upper Bound	Std. Error	RSE (%)
10	238.36	234.61	242.11	18.88	7.92
15	238.08	235.20	240.96	14.51	6.10
20	239.86	237.29	242.43	12.94	5.40
25	238.75	243.35	241.05	11.60	4.81
30	241.07	238.95	243.18	10.67	4.43
35	241.39	239.48	243.31	9.65	4.0

DISCUSSION AND CONCLUSION

This study presents a simple statistical analysis of the data taken from 5 a hectare stand using systematic sampling. The stand estimates was obtained using SRS estimators. It can be noticed from the results that actual stand estimates (original) have smaller relative standard error as compared to both estimates of trees that were marked for cutting and the trees that would be left. In terms of precision of estimates, variables under "original trees" are more precise. This is followed by the variables under trees that are to be remained/left and lastly the variables "trees to be cut". This can be explained by the fact that y_i is substantially reduced especially for trees that were marked for cutting. Take note that out of 264 trees observed only 33 trees were marked for cutting. Also, some plots have

missing observations (no trees marked for cutting) and yet they were still considered in the analysis using the same sample size (22). The same trend was observed either analyzing the data on a plot basis ($n=22$ plots) or from the total number of elements observed (here n is no longer 22 but say, 264 elements). Though the estimates using the latter approach yielded lower % SE as compared to the estimates on a per plot basis, the result is still not valid considering that we are changing our sample size. Also, we are now considering the individual trees as our observation units rather than the individual plots themselves.

The variability or dispersion of individual values in the population or the individual estimates in each plot greatly affected the %SE and hence the precision of the estimates. This is shown by the results of the variables under "trees to be cut". There were 4 plots having no trees marked for cutting which means that each of these 4 plots have 0 values (estimates). This resulted to a greater variation and therefore resulted to higher relative standard error.

From the simulation study, it indicated an inverse relationship between %SE and sample size. The results revealed that when using bigger sample size there is a reduction in the relative standard error which means that sampling estimates becomes more efficient. This means that when n approaches to N , s_y becomes zero (Kleinn, 2005).

From this study, it can be concluded that the variability of the estimates were affected by sample size and the variability of individual units in the population or the individual estimates (in this case, estimates in each plot).

LITERATURE CITED

- FREESE, FRANK**, 1962. Elementary Forest Sampling. USDA Agricultural Handbook No. 232. USDA Department of Agriculture. Forest Service. Reprinted 1971.
- KLEINN, C.** 2005. Lecture Notes for the Teaching Module Forest Inventory. Draft Version. Faculty of Forest Sciences and Forest Ecology, Göttingen, Germany.
- LOETSCH F. and K. E. HALLER**, 1973. Forest Inventory. 1 (2): BLV Verlagsgesellschaft.
- SCHREUDER, H.T., T.G. GREGORIE, and G. B. WOOD**, 1993. Sampling Methods for Multiresource Forest Inventory. John Wiley & Sons, Inc.
- SHIVER, B. D. and B .E. BORDERS**. 1996. Sampling Techniques for Forest Resource Inventory. John Wiley & Sons, Inc.

ACKNOWLEDGMENT

The author wishes to acknowledge Prof. Dr. Cristoph Kleinn and Dr. Lutz Ferhman for their help in the realization of this research project. Also, I would like to thank Mr. Christian Kleinschmit for the production of the map.