Silva Fennica 1991, Vol. 25 N:o 2: 77–83

The evaluation of forest inventory designs using correlation functions

Kari T. Korhonen & Matti Maltamo

TIIVISTELMÄ: KORRELAATIOFUNKTIOT METSÄNINVENTOINNIN OTANNAN TEHOKKUUDEN VERTAILEMISESSA

Korhonen, K.T. & Maltamo, M. 1991. The evaluation of forest inventory designs using correlation functions. Tiivistelmä: Korrelaatiofunktiot metsäninventoinnin otannan tehokkuuden vertailemisessa. Silva Fennica 25(2): 77–83.

Correlation functions of the mean volume, land use class and soil class were estimated using the data of the Finnish National Forest Inventory. Estimated functions were used for approximating the standard error of e.g. the mean volume of a cluster of plots. Standard error estimates can be used for comparing different inventory designs.

Tutkimuksessa esitetään valtakunnan metsien inventoinnin aineistosta lasketut keskitilavuuden, metsämaan osuuden ja maaluokkien osuuden korrelogrammit. Korrelogrammeihin sovitetttiin yhtälöt epälineaarisella regressiolla. Sovitetut yhtälöt antavat estimaatin korrelaatiolle koealojen välisen etäisyyden funktiona mainituille muuttujille. Korrelaatiofunktioiden avulla laskettiin mm. lohkon keskitilavuuden estimaatin keskivirheitä erilaisilla lohkomuodoilla. Esitetyn menetelmän avulla voidaan vertailla erilaisten ja erikokoisten lohkojen tehokkuutta maastotietoon perustuvassa inventoinnissa.

Keywords: forest inventories, correlation analysis, correlograms, spatial correlation, sampling error.
FDC 524

Authors' addresses: *Korhonen:* The Finnish Forest Research Institute, Joensuu Research Station, P.O. Box 68, SF-80101 Joensuu, Finland; *Maltamo*: University of Joensuu, Faculty of Forestry, P.O. Box 111, SF-80101 Joensuu, Finland.

Accepted September 5, 1991

1 Introduction

Systematic sampling designs are widely used in large area forest inventories. Gathering plots into clusters (often called tracts) has been found to be the most efficient method in most forest populations. Choosing the optimal size of sample plots, optimal distance between plots, and optimal shape and size of clusters requires information on

1) the time consumption of different working phases, and
2) the variation of the population.

The first goal of this paper is to study the 'long distance' variation of the Finnish forests. The variation is described with correlation functions. The second goal of this study is to use these functions for comparing the relative reliability of results obtained with different sampling designs. The methods used and described in this paper are similar to those used in Sweden when the new national forest survey was planned (see e.g. Ranneby et al. 1987).

2 Material

Sample plots measured in the seventh National Forest Inventory of Finland (later called NFI7 in this paper) are used to study the variation of some variables in the Finnish forests. Sample plots are located on clusters, in the shape of a half-square. The distance between plots (= relascope points) within a cluster is 200 m and the distance between clusters is 8 km. Each cluster consists of 21 sample plots (see Fig. 1) (Kuusela and Salminen 1969).

Several characteristics describing the site and growing stock on the plots are registered. A relascope (factor 2) is used for choosing the trees to be measured. In addition to these 21 so-called tally plots there are 20 so-called stump plots in each cluster. In the 'stump plots' only stumps and some stand characteristics are registered. These plots were not used in the study.

In Lapland, the NFI7 was based on both aerial photographs and ground data. For this reason, fewer field sample plots were measured in Lapland than in southern Finland. Thus, sample plots measured in Lapland were not used in this study. The study area is presented in Fig. 2.

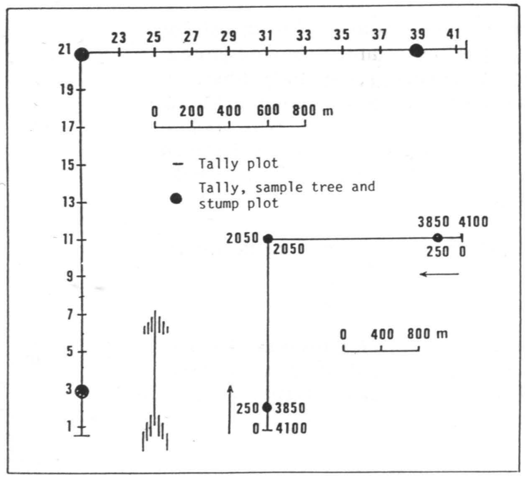

Figure 1. The tract scheme of the Finnish NFI.

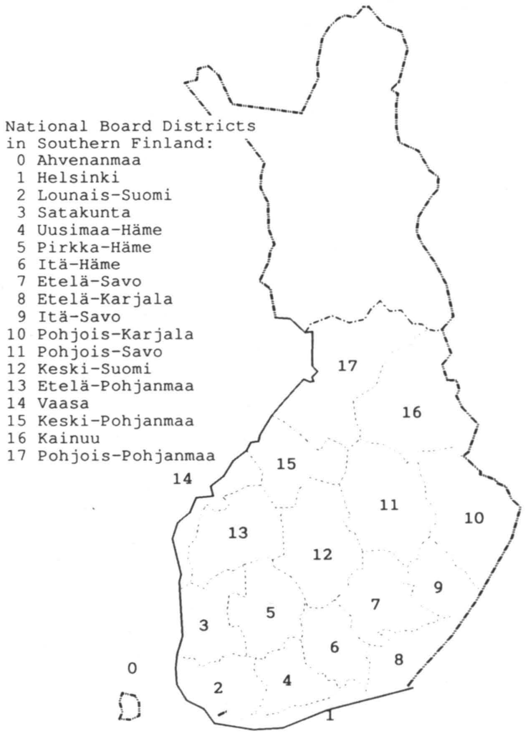

Figure 2. The study area.

3 Methods

3.1 Estimation of correlograms and correlation functions

Correlograms were calculated for mean volume on forest land, land use class (forest/non forest land) and soil class (mineral soil/peatland). Mean volumes (m^3/ha) were estimated for each plot by summing the volumes per hectare of the measured trees. The land use class (forest or non-forest) and soil class (mineral soil or peatland) of the plot were determined according to it's middle point.

Correlograms were estimated by calculating the variances (σ^2) and variograms for above-mentioned variables. Variograms were calculated with the following formula (see e.g. Ranneby 1981).

$$\hat{v}(u) = \sum_{i=1}^{n} (X(s_i + u) - X(s_i))^2 \qquad (1)$$

where $X(s_i)$ = value of variable x at point s_i,
u = distance, and
n = number of points (= sample plots).

The correlation of, e.g., the mean volumes of plots as a function of distance between plots was estimated using formula (2).

$$\hat{c}(u) = 1 - \hat{v}(u)/s^2, \qquad (2)$$

where s^2 estimates σ^2.

The observed correlations were smoothed into continuous functions with non-linear regression analysis.

3.2 Estimation of sampling error with correlation functions

When the correlation between clusters is zero (= clusters are far from each other) the relative reliability of an inventory design can be assessed by the error variance of a cluster mean (Matern 1960). Estimated correlation functions were used for estimating the error variance, e.g., for the mean volume estimate obtained with one cluster.

The cluster mean is calculated with formula:

$$\bar{x} = 1/n \cdot \sum_{i=1}^{n} x_i \qquad (3)$$

The formula for estimating the error variance is:

$$\mathrm{var}(\bar{x}) = 1/n^2 \cdot \mathrm{var} \sum_{i=1}^{n} x_i$$

$$= 1/n \cdot \mathrm{var}(x) + 2/n^2 \cdot \sum_{i \neq j} \mathrm{cov}(x_i, x_j). \qquad (4)$$

In order to use formula (4) we need
1) the variance of variable x on plot i, and
2) the covariance of values of variable x as a function of the distance between plots.

The expected value for the variance is assumed to be equal on every plot and was estimated from the NFI7 data. The standard deviations of mean volumes of regions I, II, and III were 91.3, 97.4, and 72.1 m^3/ha, respectively. The standard deviations of percentage of forest land were 0.48, 0.39, and 0.46, respectively.

The covariances were obtained from correlation functions (see chapter 31) using the following formula.

$$\mathrm{cov}(y_1, y_2) = s(y_1) \cdot s(y_2) \cdot \mathrm{corr}(y_1, y_2), \qquad (5)$$

where $s(y_i)$ = standard deviation of variable y_i, and
$\mathrm{corr}(y_1, y_2)$ = correlation of variables y_1 and y_2.

In this case $s(y_1) = s(y_2) = s(x)$. Thus, we obtain:

$$\mathrm{cov}(x_i, x_j) = \mathrm{var}(x) \cdot \mathrm{corr}(x_i, x_j),\ i,j = 1,n. \qquad (6)$$

4 Results

4.1 Correlograms

At first, correlograms were calculated for each National Board district. Correlograms were also calculated separately for North-South -direction and East-West -directions.

We observed that the variation was almost isotropic, i.e. the correlations calculated in the North-South -direction were about equal to those calculated in the East-West -direction. Thus, the final correlograms were calculated for combined data. Furthermore, data from different National Board districts were combined: Finland was divided into three separate groups of National Board districts according to the pattern of variation. The groups were (see Fig. 2):

I = South-east and South-west Finland (National Board districts 1–6 and 8)
II = Central and Eastern Finland (districts 7, 9–12), and
III = Ostrobothnia (districts 13–17).

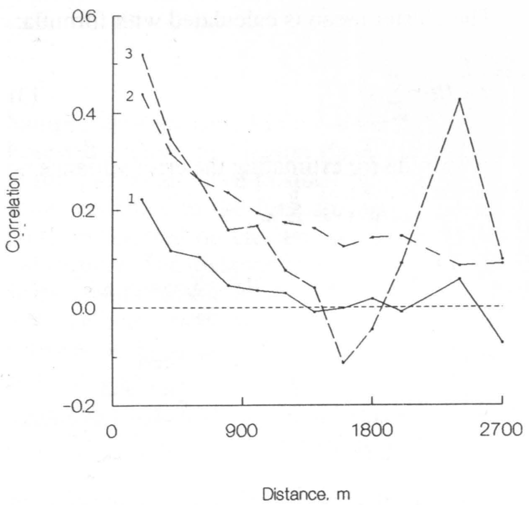

Figure 3. Estimated correlations of (1) mean volume, (2) land use class, and (3) soil class. Region I.

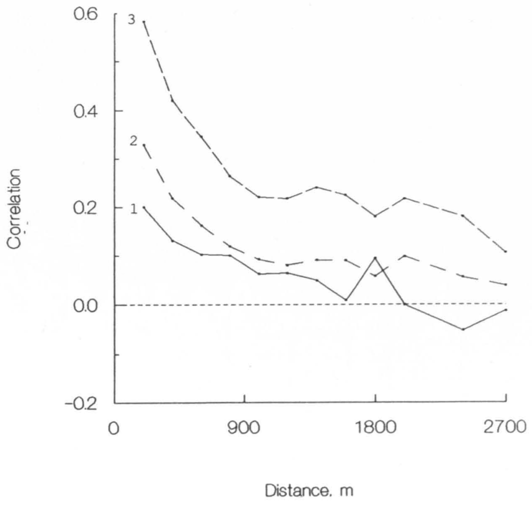

Figure 4. Estimated correlations of (1) mean volume, (2) land use class, and (3) soil class. Region II.

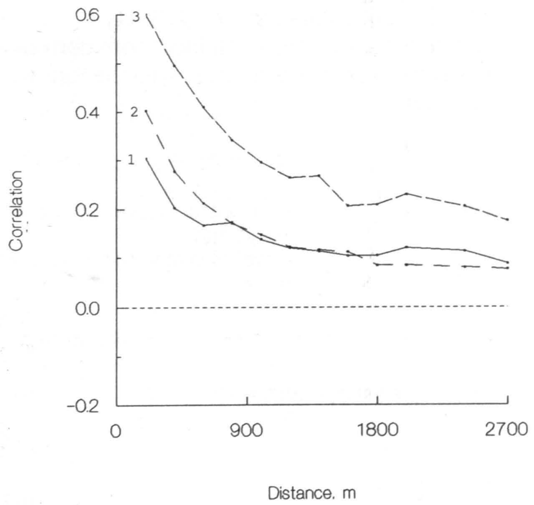

Figure 5. Estimated correlations of (1) mean volume, (2) land use class, and (3) soil class. Region III.

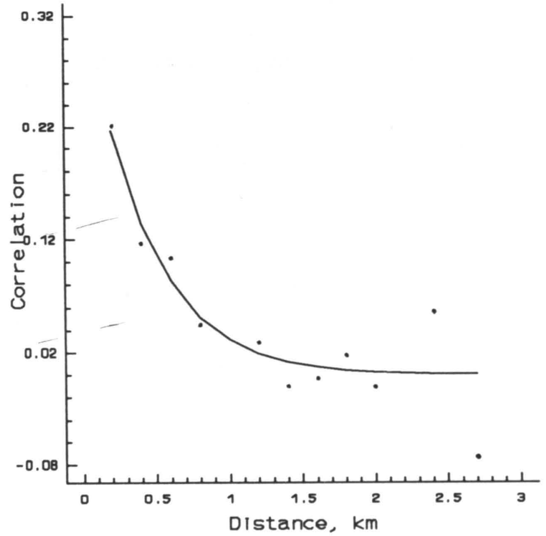

Figure 6. Estimated correlation function for the mean volume of region I.

Correlograms for different regions and for different variables are presented in Figs. 3–5. Correlations from 200 m to 2 km were calculated using plots on the same side of the half-square cluster. Correlations at the distance of 2.3 km were calculated using plots numbered 3 and 37 or 5 and 39 (see Fig. 2). At the distance of 2.7 km plots 3 and 41 or 1 and 39 were used.

Correlograms (Figs. 3–5) show, e.g., that the mean volume of sample plots in region III is correlated at longer distances than in the other two regions. The land use class in region I has a spatially greater correlation than in the other regions. This is probably due to the fact that there are larger agricultural fields in the southern Finland than in the other parts of the country.

4.2 Correlation functions

Non-linear regression was applied to fit functions into the correlograms of the mean volume and land use class. Several functions, e.g. logarithmic, exponential and power functions, were tested. In regions I and II the best function for describing the mean volume correlogram was found to be:

$$c(u) = a_0 \cdot e^{-a_1 u}, \qquad (7)$$

where a_0 and a_1 are parameters, and
u = distance between plots in kilometers.

The best fit for the mean volume correlogram in region III and the correlograms of land use class in all regions was obtained with function (8).

$$c(u) = a_0 \cdot u^{-a_1}. \qquad (8)$$

Figure 6 gives an example of a fitted correlation function. Parameter estimates for different regions and different variables are presented in Table 1.

It would be logical that the estimated correlation functions would give estimate 1 for distance 0. However, this logical constraint was ignored for two reasons:

1) in the applications correlation estimates are not needed for distances less than 100 meters, and
2) in the applications most accurate estimates for correlations are needed for distances higher than 200 m.

4.3 Comparison of different sampling designs

A Fortran-program was written to calculate error variances of the estimates obtained using clusters of different sizes and shapes. The program uses formulas (6), (7), and (8) to estimate the covariances between plots within a cluster. The error variances of the mean volume and forest land percentage estimates are estimated with formula (4).

Table 1. Parameter estimates for the correlation functions of mean volume and land use class.

Region	Variable	Function	$\hat{a}_0$	$\hat{a}_1$
I	mean volume	(7)	0.352	2.406
II	mean volume	(7)	0.250	1.316
III	mean volume	(8)	0.141	0.460
I	land use class	(8)	0.189	0.540
II	land use class	(8)	0.108	0.670
III	land use class	(8)	0.145	0.649

Table 2 shows some results obtained with half-square clusters of different sizes and varying distances between plots. In Table 2 the number of plots per cluster varies from 17 to 25 and the distance between plots from 100 m to 400 m. The cluster of 21 plots with the distance of 200 m (6th line in Table 2) is the cluster pattern used e.g. in the 7th and 8th National Forest Inventory of Finland. Table 2 shows, as expected, that the error variances decrease when the distance between plots increases. It might seem somewhat surprising that the error variance of the mean volume estimates is higher at region II than at region III even though the mean volumes are more correlated in the latter region. This is explained by the fact that the variance of the mean volume is lowest at region III.

Table 3 shows results obtained with square-shaped clusters. Comparison of Tables 2 and 3 shows that with equal distance between plots and about equal number of plot per cluster the half-square shaped clusters give lower error variances than square shaped clusters. This is due to the fact that in square shaped clusters the average distance between plots is lower than in half-square shaped clusters (see e.g. Matern 1960).

Table 2. Error variances of mean volume (m^3/ha) and forest land percentage estimates obtained with different half-square clusters. Notation: Dist. = distance between plots, No = number of sample plots/cluster, Volume = mean volume, Forest % = percentage of forest land.

		Error variances					
		Region I		Region II		Region III	
Dist	No	Volume	Forest %	Volume	Forest %	Volume	Forest %
100	17	865.5	0.040	1047.7	0.020	723.1	0.033
200	17	641.0	0.032	834.1	0.016	609.2	0.026
300	17	563.1	0.028	723.3	0.014	557.6	0.023
400	17	531.0	0.026	662.9	0.013	526.4	0.021
100	21	701.4	0.036	892.0	0.018	642.8	0.029
200	21	503.4	0.028	671.8	0.014	534.9	0.022
300	21	445.7	0.025	573.0	0.012	486.1	0.019
400	21	423.8	0.023	524.8	0.011	456.5	0.018
100	25	579.9	0.032	770.2	0.015	583.2	0.026
200	25	411.2	0.025	554.9	0.012	480.8	0.020
300	25	368.1	0.022	470.1	0.011	434.4	0.017
400	25	352.4	0.020	432.3	0.009	406.3	0.016

Table 3. Error variances of mean volume (m^3/ha) and forest land percentage estimates obtained with different square clusters. Notation: Dist. = distance between plots, No = number of sample plots/cluster, Volume = mean volume, Forest % = percentage of forest land.

		Error variances					
		Region I		Region II		Region III	
Dist	No	Volume	Forest %	Volume	Forest %	Volume	Forest %
100	16	971.2	0.042	1090.7	0.022	742.0	0.036
200	16	759.6	0.033	934.7	0.017	628.2	0.028
300	16	657.4	0.029	833.1	0.016	576.6	0.024
400	16	604.4	0.027	765.7	0.015	545.4	0.022
100	20	854.3	0.040	1006.2	0.020	704.6	0.034
200	20	613.0	0.031	800.5	0.016	583.2	0.025
300	20	517.3	0.027	682.6	0.014	528.2	0.022
400	20	474.4	0.025	612.9	0.013	495.0	0.020
100	24	747.6	0.038	927.1	0.019	671.8	0.031
200	24	503.0	0.029	688.7	0.014	547.6	0.023
300	24	421.1	0.025	567.7	0.013	491.3	0.020
400	24	388.2	0.023	503.2	0.011	457.2	0.018

5 Discussion

The methods presented in this paper can be used to compare the precision of different inventory designs. In order to compare the efficiency, the cost of the inventory is also needed (see e.g. Ek et al. 1984). If measurements on a plot and data processing costs are assumed to be fixed, the cost of a field inventory can be determined according to the time consumption of the field work.

It is believed that in Finnish conditions the most efficient way, by far, is to use such clusters that can be measured during one working day. If this assumption is true the effect of the distance between plots and the shape of a cluster on the efficiency of an inventory can be studied simply by comparing only those clusters that can be measured during one working day. If the distance between plots is e.g. 200 m, more plots can be located on a cluster than with a distance of 300 m between plots. Measuring square shaped clusters takes less time than measuring half-square shaped clusters (see e.g. Matern 1960). Thus, more sample plots can be located on square clusters than on half-square clusters.

The correlograms and correlation functions presented in this paper are based on plots measured with a relascope with basal area factor two. If larger sample plots are used, the results in this paper are no longer valid.

This paper presents results concerning the estimation of mean volume and land use class only. Usually, these are not the only goals of a forest inventory — we might be interested as well in the health or recreational values of the forests. Thus, when the inventory design is chosen, several variables should be taken into consideration. This can be laborious, but it is better to have imperfect information to support the decisions than no information at all.

References

Ek, A.R., Scott, C.T., Zeisler, T.R. & Benessallah, D. 1984. Cluster sampling: from theory to practise. In: Inventorying forest and other vegetation of the high latitude and high altitude regions: Proceedings of the conference; Fairbanks, Alaska, Society of American Foresters, 84-11. p. 78–81.

Kuusela, K. & Salminen, S. 1969. The 5th National Forest Inventory in Finland. Communicationes Instituti Forestalis Fenniae 61(4). 72 p.

Matern, B. 1960. Spatial variation. Stochastic models and their application to some problems in forest surveys and other sampling investigations. Meddelanden från Statens Skogsforskningsinstitut. Band 46, Nr 5. 144 p.

Ranneby, B. 1981. Den topografiska variationen inom olika områden. En redovisning av skattade korrelationsfunktioner. Sveriges Lantbruksuniversitet, Institutionen för skogstaxering. Projekt NYTAX 83. Rapport 5. 54 p.

— , Cruse, T., Hägglund, B., Jonasson, H. & Svärd, J. 1987. Designing a new national forest survey for Sweden. Studia Suecica 177.

Total of 5 references