SILVA FENNICA

Vol. 28(1), 1994

SILVA FENNICA
a quarterly journal of forest science

Publishers The Finnish Society of Forest Science
The Finnish Forest Research Institute

Editors Editor-in-chief Eeva Korpilahti
Production editors Tommi Salonen, Seppo Oja

Editorial Office Unioninkatu 40 A, FIN-00170 Helsinki, Finland
Phone +358 0 857 051, Fax +358 0 625 308, E-mail silva.fennica@metla.fi

Aim and Scope Silva Fennica publishes original research articles, critical review articles, research notes reporting preliminary or tentative results, and discussion papers. The journal covers all aspects of forest research, both basic and applied subjects. The scope includes forest environment and silviculture, physiology, ecology, soil science, entomology, pathology, and genetics related to forests, forest operations and techniques, inventory, growth, yield, quantitative and management sciences, forest products, as well as forestry-related social, economic, information and policy sciences.

SILVA FENNICA

a quarterly journal of forest science

Vol. 28(1) · 1994

The Finnish Society of Forest Science
The Finnish Forest Research Institute

New Scope of Silva Fennica

In today's world, forests continue to be a natural resource essential to many individual nations and vital to all mankind. Research is required to increase our knowledge of this resource, of issues such as ecological interactions and variety of forest ecosystems, silvicultural management, sustainable utilization of natural resources, technical development and facilities, and social and cultural values of forests and forestry. Many problems are encountered and shared by scientists working in different countries, while other problems are specific to particular areas but equally important.

For decades, it has been the policy of Silva Fennica to concentrate on research articles relevant to Finnish forestry. In recent years, the journal has published a growing number of submissions dealing with forestry elsewhere in the boreal zone, and even in the more southerly conditions of developing countries. In view of the above considerations and the growing interaction within the scientific community, the publishers have now decided that Silva Fennica will start publishing papers over the entire range of forest science. With the new international dimension, Silva Fennica will be a forum for dissemination of research results and exchange of ideas among the forest research community in its widest sense. Manuscripts of original research articles and constructive reviews are welcome from authors in all countries.

It is my intention to maintain and improve the scientific standard of the journal and arouse ever greater interest in its contents. I aim at a competent and constructive review of manuscripts. In this endeavour, I am happy to be supported by the expertise of the newly appointed international editorial board consisting of recognized experts in various fields of forest research. I also strive to preserve Silva Fennica's record of speedy printing: the average time for a paper to appear in print has been about four months from acceptance.

I expect that the renewed Silva Fennica will be able to cater interesting research results and stimulating discussion papers to all concerned. I would also like to call the attention of scientists worldwide to the new scope of the journal, hoping that Silva Fennica will prove able to serve their needs for a publication forum.

Eeva Korpilahti
Editor-in-Chief

Classical and Model Based Estimators for Forest Inventory

Annika Kangas

Kangas, A. 1994. Classical and model based estimators for forest inventory. Silva Fennica 28(1): 3–14.

In this study, model based and design based inference methods are used for estimating mean volume and its standard error for systematic cluster sampling. Results obtained with models are compared to results obtained with classical methods. The data are from the Finnish National Forest Inventory. The variation of volume in ten forestry board districts in southern Finland is studied. The variation is divided into two components: trend and correlated random errors. The effect of the trend and the covariance structure on the obtained mean volume and standard error estimates is discussed. The larger the coefficient of determination of the trend model, the smaller the model based estimates of standard error, when compared to classical estimates. On the other hand, the wider the range and level of autocorrelation between the sample plots, the larger the model based estimates of standard error.

Keywords systematic cluster sampling, covariance structure, models, estimation, forest inventories.
Author's address Faculty of Forestry, University of Joensuu, P.O. Box 111, FIN-80101 Joensuu, Finland.
Accepted February 15, 1994

1 Introduction

The concept of superpopulation models was first introduced into sampling theory by Cochran (1946). In the last two decades, new methods based on superpopulation models have been increasingly studied in different fields. In the literature on forest inventory methods, hardly any attention had been paid to superpopulation models, apart from in the studies of Matérn (1960), until recently.

In recent years, the possibilities of new methods have been noted in forestry, too. Model based estimation allows the combination of different sources of information or sources of error. Mandallaz (1991) used the GIS to provide the auxiliary information needed in the superpopulation models. Information from previous surveys was used as prior information of forest inventory in Kangas (1991).

In model based estimation, the characteristics of a population are described with a model. Inference about the population depends on the assumed model, not on sampling design, as in classical methods. If the sampling units can be assumed to be independent, variation in the population can be described by a simple regression model. If sampling units are correlated, these correlations have to be given consideration in model assumptions. If the population is properly described with the model, the same model based estimators can be used for any sampling design, for complex sampling designs, and even for samples which are not selected by any objective method.

Measurement errors and errors due to volume and height models can also be taken into account when model based inference theory is used (Kangas 1993). Also, the inventory design does not have to be as precisely determined as in classical methods: additional sample plots can be measured after the initial inventory, or the inventory design can be changed without any major problems.

If there is a trend in the population, part of the variation of y is due to the trend. This situation can be presented by means of a superpolation model as

$$y_i = \alpha + \beta x_i + e_i \qquad (1)$$

where

$$E(e_i) = 0 \quad \text{and} \quad E(e_i^2) = \sigma^2 \qquad (2)$$

The errors may be either correlated or non-correlated. In (1) x_i is a variable (or a vector of variables) whose value is known for each sampling unit in the population.

In model based methods, the sampling error is estimated from the error terms of the superpopulation model. Thus, the part of the variation due to the trend is first removed. In the presence of the trend, the precision of estimators is higher, because the trend component reduces the variance of the model.

In the case of autocorrelated sampling units, the precision of systematic sampling can be higher than that of simple random sampling (SRS). This is the case if the correlation between sample points is positive, decreasing with increasing distance and convex, and, if the sample is an equilateral lattice (Bellhouse 1988, p. 134, Ranneby et al. 1987). Cochran (1977, p. 220) has given three examples of correlation functions which satisfy these conditions. These are linear, exponential, and hyperbolic tangent. Any linear combination of these functions can also be used. If the sampling design is a cluster design, systematic cluster sampling is more precise than random cluster sampling, under the same conditions.

The theory of systematic sampling has been one of the most problematic parts of classical sampling theory. Classical methods require that each sampling unit has a known, positive probability of selection. In systematic sampling the whole sample is determined, when the first sampling unit is selected and thus, this assumption does not hold. The most difficult thing in systematic sampling designs, with respect to classical sampling theory, is that no general estimators of sampling errors can be found. The standard error can be presented with a general formula (e.g. Ripley 1981), but this formula cannot be used in most practical situations, since the autocorrelation function required in the formula is usually not known.

In forest inventory, systematic sampling methods have been used since the beginning of this century. Many formulas have been proposed for estimating standard error in systematic sampling. The estimators that have proved to be useful in practice are based on quadratic forms or on spatial smoothing of interesting variables (Matérn 1947). In these estimators, the part of variation due to the trend is removed, in the same way as in model based methods. The formulas of Lindeberg (1924) and Langsaeter (1926–27) for line surveys are pioneer work in this field. The formulas for plot surveys were first investigated in Matérn (1947, 1960) and later in Ranneby (1981a).

In model based methods, the estimated model is also used for estimating the population mean. In classical methods, the trend component is usually not considered when the mean values are calculated. The trend can, however, also be taken into account in formulating the classical estimators of mean (Bellhouse 1988, p. 130). For example, the mean of the smallest and largest values in the sample can be used in the estimation of population mean, in addition to the ordinary sample mean.

The formulas presented first by Matérn (1947, 1960) and further developed by Ranneby (1981a), have been used in the Finnish National Forest Inventory for as long as the systematic cluster design has been in use. Yet, the covariance structure in different parts of Finnish forests and its effect on estimators has not been thoroughly studied. Different standard error estimators were compared by Päivinen (1987), but the effects of variation in the covariance structure were not studied. The effects of distance between sample plots within the clusters, on the standard errors have been studied by Korhonen and Maltamo (1991). In this study southern Finland was divided into three parts in which the covariance structure was studied separately.

The aim of this study is to compare different model based estimators of mean and standard error with classical estimators in different conditions. The classical estimates of standard error are calculated by quadratic forms. The covariance structure is described by a trend and covariance functions. These functions are estimated for ten forestry board districts, which are different with respect to size, and to other characteristics. The effect of covariance structure on estimators is discussed.

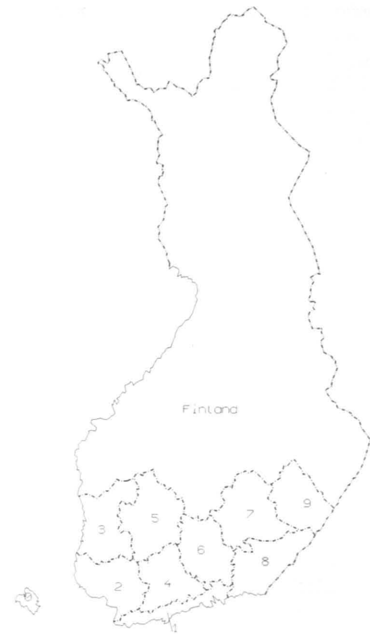

Fig. 1. The study area.

2 Material and Methods

2.1 Material

The material used in the calculations is from the VIII National Forest Inventory of Finland, from ten forestry board districts in southern Finland. The study area is presented in Fig. 1. The measurements took place in 1986–1989. The sample plots were measured in clusters, which are called tracts. The shape of the tract is a half-square (Fig. 2). The distance between tracts is 8 km in north-south direction and 7 km in west-east direction. The distance between sample plots within the tracts is 200 m.

There are 21 sample plots in each tract, three of which are sample tree plots. In sample tree plots, the height and upper diameter of each tree

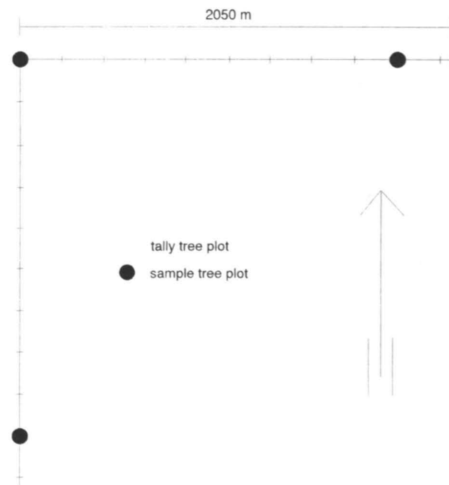

Fig. 2. The tract scheme.

is known. In the tally tree plots, only diameters were measured, and the height of the trees is estimated using height models (Henttonen, unpubl.). For each tree, the volume is estimated using volume models (Laasasenaho 1982) and then the volume per hectare on the plot is calculated.

2.2 Models Describing the Covariance Structure

The model used for describing the variation of volume can generally be presented as (e.g. Mandallaz 1991)

$$y_i = X_i\beta + \varepsilon_i \tag{3}$$

In this model the surface is divided into large-scale variation (X'ß) and small-scale random process. The vector X_i contains the values of additional variables at point i. The values of these variables have to be known at each point of the study area.

It would have been possible to use a model without the trend component as

$$y_i = \mu + \varepsilon_i \tag{4}$$

but the standard error of the model can be reduced if the trend component is included. Also, the range of autocorrelation can be reduced with a trend component. In many cases, making the difference between the trend component and random, correlated variation is a matter of personal consideration (Cressie 1986). In this study, the reduced model was used for testing the significance of independent variables in model (3).

The fitted trend of mean volume was a quadratic surface

$$\begin{aligned} y_i = {} & \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{1i}^2 + \beta_5 x_{2i}^2 \\ & + \beta_6 x_{3i}^2 + \beta_7 x_{1i} x_{2i} + \beta_8 x_{1i} x_{3i} + \beta_9 x_{2i} x_{3i} + \varepsilon_i \end{aligned} \tag{5}$$

where x_1,x_2 are the coordinates and x_3 is the altitude.

For a two-dimensional model, the full quadratic surface should be estimated, although all of the parameters are not statistically significant. In this way, the surface is invariant under the rigid motions of the coordinate scheme (Ripley 1981, p. 35). In this study, the full model was estimated for the three-dimensional case, although, for the altitude, all the terms might not have been necessary.

If the process is assumed to be isotropic, the covariance functions have to satisfy certain restrictive conditions besides the positive-definiteness condition. Thus, it is not possible to use just any function as a covariance function. In this study, Whittle's function (Ripley 1981)

$$C(r) = \frac{\sigma^2}{2^{v-1}\Gamma(v)}\left(\frac{r}{r_0}\right)^v K_v\left(\frac{r}{r_0}\right) \tag{6}$$

was used, because of its flexibility in short distances.

In (6) the term K_v is a modified Bessel function of second kind and order v, and Γ is the gamma function. The parameters to be estimated are r_0(km) and v. Parameter r_0 describes the range of correlation (Fig. 3a) and parameter v describes the level of correlation (Fig. 3b).

The parameters of the trend model and covariance functions were estimated by the maximum likelihood method (Cook and Pocock 1983). The ML estimation for a particular parametrization of correlation matrix R ($\Sigma = \sigma^2 R$) and $\varepsilon_{ji} \sim N(0,\sigma^2)$, amounts to finding $R = \hat{R}$, which minimizes

$$g(\hat{R}) = \log|\hat{R}| + N \log \hat{\sigma}^2 \tag{7}$$

where

$$\hat{\sigma}^2 = \frac{1}{N}(Y - X\hat{\beta})' \hat{R}^{-1}(Y - X\hat{\beta}) \tag{8}$$

and

$$\hat{\beta} = (X'\hat{R}^{-1}X)^{-1} X'\hat{R}^{-1}Y \tag{9}$$

The estimation was carried out using the grid search method. The step size in the grid search for parameter r_0 was 0.5 and for parameter v 0.005.

The variance-covariance matrices were very large, which makes inverting of the matrices quite difficult. Thus, it had to be assumed that

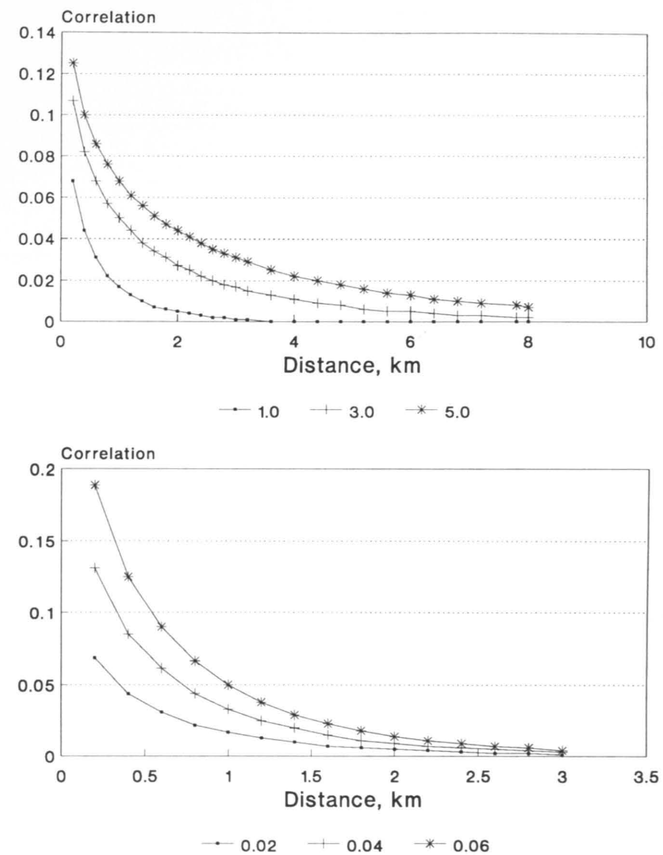

Fig. 3. (a) The Whittle correlation function with v = 0.02 and different values of r_0. (b) The Whittle correlation function with r_0 = 1.0 and different values of v.

correlation between the sample plots diminishes rapidly. Calculations were made according to the assumption that the sample plots in each tract are correlated with the plots in surrounding eight tracts. In Kangas (1993), correlations were not restricted, and the results obtained support this assumption. In this case, the variance-covariance matrix is band-diagonal, although the band is quite broad. With this assumption, it was possible, although very time-consuming, to calculate the parameters of covariance matrices even for districts with 3 200 sample plots. The calculations were made using IMSL procedures with a VAX-VMS computer.

The Cholesky decomposition of correlation matrix R, R = LL', was utilized in parameter estimation (Ripley 1981). The equation (9) was modified into form

$$\hat{\beta} = ((L^{-1}X)'(L^{-1}X))^{-1}(L^{-1}X)'(L^{-1}Y) \tag{10}$$

in order to simplify the calculations.

The significance of the trend surface was calculated using the error sum of squares of the reduced model (SSE, eq. 4) and the full model (SSE_f, eq. 3) as follows:

$$F = \frac{(SSE - SSE_f)/(p_f - p)}{SSE_f/(n - p_f)} \qquad (11)$$

In (11) p is the number of predictors in the reduced model and p_f is the number of predictors in the full model. The R^2 was calculated using the same error sum of squares as (Kvålseth 1985)

$$R^2 = 1 - \frac{SSE_f}{SSE} \qquad (12)$$

and the adjusted R^2 by dividing the error sums of squares in (12) with their degrees of freedom, $(n-p_f)$ and $(n-p)$, respectively.

A simpler assumption is a mixed model

$$y_{ji} = X_{ji}\beta + c_j + e_{ji} \qquad (13)$$

where $c_j \sim N(0,\sigma_c^2)$ is the random cluster effect and $e_{ji} \sim N(0,\sigma_e^2)$ is the random plot effect, and c and e are independent. This model means that a constant correlation

$$\rho = \frac{\sigma_c^2}{\sigma_c^2 + \sigma_e^2} \qquad (14)$$

is assumed within a cluster, and clusters are assumed to be mutually independent. Estimation of random effects is much easier than estimation of covariance functions, which makes this model a desirable alternative to the general model. Variances of the random effects were estimated using Henderson's fitting constant method (Searle 1971).

2.3 Methods for Estimating the Mean Volume and Standard Error

In classical methods, the estimate of population mean for systematic sample design is usually obtained by sample mean as in simple random sampling. Because some of the sample plots are not totally on forest land, the mean volume in the Finnish NFI is obtained by weighing the sample plots in proportion to the percentage of forest land of the total area of sample plots.

The standard error of mean volume (on forest land) in the Finnish NFI is estimated with the quadratic form (Päivinen 1987; Salminen 1973)

$$\mathrm{var}(\hat{y}) = \frac{q\sum_{i=1}^{m} T_i}{n^2} \qquad (15)$$

where
$\hat{y}$ = the estimate of mean volume,
n = total number of sample plots,
m = total number of tract groups,
q = number of tracts each tract group represents

and

$$T_i = \frac{1}{4}(z_{i1} - z_{i2} - z_{i3} + z_{i4})^2 \qquad (16)$$

and

$$z_{ij} = (y_j - y)n_j, \qquad (17)$$

where
y_j = mean volume on forest land in cluster j (tract j),
y = sample mean volume on the area and
n_j = number of sample plots on forest land in the tract j.

One tract group consists of four tracts (Fig. 4). In this study the tract groups are formed in such a way that each tract belongs to four groups. Thus, each tract group represents 1 tract, i.e. the value of q is one. In Fig. 4 the two tract groups are formed so that each tract belongs to one tract group and the value of q thus is 4.

If the correlation decreases slowly as a function of distance, this quadratic form gives overestimates of standard error (Matérn 1960). Also, the quadratic form has quite a large variance (Ranneby 1981a), which may affect the results for small areas. It has been recommended not to use this formula in calculating variance estimates for small areas. There should be at least 30, preferably 100, tract groups in the area, in order to ensure good estimates. (Salminen 1985).

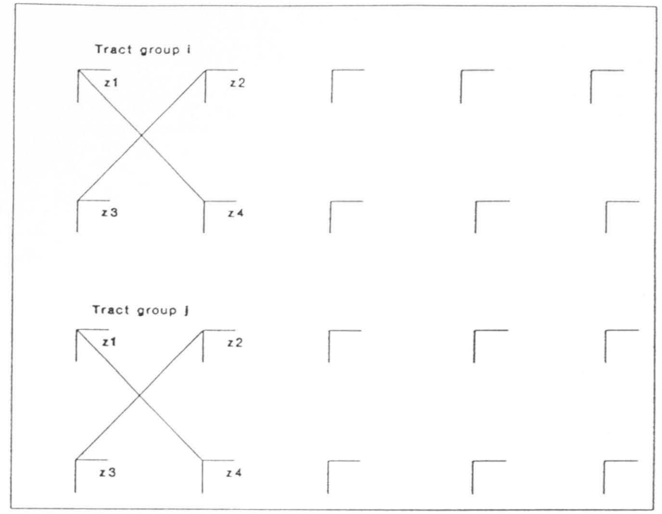

Fig. 4. Formation of two tract groups for calculation of standard error. The two tract group represent 4 tracts each.

The standard error for random two-stage cluster sampling can be estimated by (Loetsch et al. 1973)

$$\mathrm{var}(\hat{y}) = \frac{\sum_{j=1}^{J} n_j(y_j - y)^2}{(\sum_{j=1}^{J} n_j)(J-1)} \qquad (18)$$

where J is the number of tracts. The other notations are as above.

The formula of two-stage cluster sampling has been used to illustrate the behaviour of model based estimators in different situations.

In model based methods, all inference about the population depends on the assumed model. The predictor of the population mean is then

$$\hat{y} = \bar{X}\hat{\beta} = \bar{X}'(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y \qquad (19)$$

and standard error of mean volume can be estimated, for example, by

$$\mathrm{var}(\hat{y}) = \bar{X}'\,\mathrm{var}(\hat{\beta})\bar{X} = \bar{X}'(X'\Sigma^{-1}X)\bar{X} \qquad (20)$$

The autocorrelations between the sample plots are taken into account in the variance-covariance matrix Σ.

If the model is biased, variance (20) can be a poor estimate of the mean square error of estimated mean volume. When the number of observations is large, the sample is representative for the population, and the sample plots are selected by an objective method, the variance estimator can be used quite safely. The sample can be considered as representative, if the sample is balanced, i.e. the sample moments are close to those of the population (Cassel et al. 1977).

3 Results

The trend component, described by a quadratic surface (5), in the general model (3) was significant, with the risk level 0.05 in all of the districts except for 0 and 9 (Table 1). In district 0, the number of sample plots is small, and this affects the degree of significance. In district 9, the trend clearly has no effect: the adjusted coefficient of

Table 1. Coefficient of determination (%), adjusted coefficient of determination (%), standard error of estimated model and F statistics for each district.

District	R^2 %	R^2_{adj} %	s_e	F	n
0	3.595	0.664	114.2	0.2781	306
1	4.095	3.537	104.3	0.0000	1550
2	3.406	2.989	95.8	0.0000	2095
3	2.730	2.399	91.0	0.0000	2657
4	1.193	0.719	105.5	0.0073	1886
5	3.304	3.041	95.8	0.0000	3010
6	1.976	1.591	97.4	0.0000	2301
7	0.776	0.493	97.6	0.0034	3173
8	2.252	1.888	97.3	0.0000	2429
9	0.474	0.000	107.1	0.4907	1783

Table 2. The parameters of covariance functions (6) and estimated within-cluster correlations (14) for each district

District	υ	r_0, km	ρ
0	0.025	1.0	0.046
1	0.060	1.0	0.085
2	0.025	5.5	0.073
3	0.040	2.0	0.085
4	0.065	1.0	0.066
5	0.060	1.0	0.058
6	0.045	1.0	0.060
7	0.050	1.0	0.042
8	0.035	1.0	0.045
9	0.050	1.0	0.076

determination is zero. The standard error of the model was also largest in these districts.

Differences in the covariance structure between the districts are mainly revealed in the variation of values of parameter ν (Table 2). In districts 0 and 2, the parameter ν has its smallest value. In other districts differences are not so clear.

The value of parameter r_0 is the same, 1.0 km, for most of the districts. This means that there is only within-cluster correlation after the effect of the trend is removed. In district 2, however, the value of r_0 was 5.5 km. This district is partly coastal region and partly inland. The model used obviously does not sufficiently decribe the trend surface in this district, which probably causes this result.

This result indicates, that the quadratic trend model is not the best choice for all of the districts: with a more flexible model, correlations between the tracts could be reduced. Thus, it would be possible to assume that the between-cluster correlations equal zero, in which case calculations would be much easier.

The mean volumes obtained with different methods are shown in Table 3. The stronger the correlation within the clusters, the nearer the mean volume obtained with the general model (3) (SYS) or mixed model (13) (MIX), the unweighted mean of cluster means (CLU). In SRS estimators, autocorrelation between sample plots is not considered, and the weight of each cluster depends on the number of forest sample plots in

Table 3. The mean volumes for each district estimated with simple random sampling (SRS), general model (3) (SYS), mixed model (13) (MIX) and cluster (CLU) estimators

District	SRS	SYS	MIX	CLU
0	154.3	152.8	150.9	141.6
1	157.8	158.5	159.1	163.3
2	141.5	141.5	142.1	147.2
3	121.2	121.5	122.1	126.2
4	154.6	155.9	155.3	158.8
5	128.1	129.3	129.1	136.2
6	144.2	144.6	144.4	146.3
7	133.8	133.7	134.0	133.5
8	127.3	127.7	127.7	129.6
9	137.1	138.0	138.1	143.1

the cluster. In cluster sampling estimators, the weight of each cluster is the same. The clusters can, however, be weighted according to their size, in which case the weight of each sample plot would be equal.

When autocorrelation between the sample plots is taken into account, the weight of single sample plots in the cluster is reduced, but the clusters with many forest sample plots still have more weight than those with few forest sample plots. However, if there were the same number of forest sample plots in each cluster, estimates obtained using different methods would be equal.

Table 4. The standard errors of mean for each district estimated with simple random sampling (SRS), quadratic form (15) (QUAD), general model (3) (SYS), mixed model (13) (MIX) and cluster (18) (CLU) estimators

District	SRS	QUAD	SYS	MIX	CLU
0	6.83	7.93	7.59	8.09	7.27
1	2.79	3.57	3.49	3.69	3.58
2	2.18	2.64	3.54	2.85	2.94
3	1.82	2.49	2.57	2.59	2.71
4	2.50	2.75	3.35	3.31	3.28
5	1.81	2.60	2.42	2.38	2.72
6	2.09	2.57	2.65	2.77	2.80
7	1.77	1.95	2.31	2.21	2.26
8	2.02	2.50	2.45	2.54	2.57
9	2.58	3.14	3.32	3.60	3.42

Table 5. Correlations between the differences in standard error estimates and R_2, parameter υ, and the standard error of the model.

	R_2	υ	s_e
SYS-QUAD	–0.78	0.53	–0.13
CLU-SYS	0.87	–0.31	0.07
CLU-QUAD	–0.60	0.60	–0.15

The larger the difference between the cluster mean and simple random sampling mean, the greater the effect of within-cluster correlations on the estimated means.

The standard errors obtained with different methods are presented in Table 4. Differences between the standard errors estimated by quadratic form (15), two-stage cluster sampling (18) and the model for systematic sampling (20), were calculated for each district. Then correlations between the differences and the characteristics of the model were calculated (Table 5). District 2 is not included in the following calculations, because the covariance structure in this district differs so much from the other districts. Although the coefficient of determination has not as straightforward an interpretation as in OLS models, the differences between the different methods can be illustrated by R^2.

The difference between the standard error of cluster sampling and systematic sampling (CLU-SYS) had the biggest correlation with R^2, 0.87 (Table 5). This could be expected, because the trend component is not at all taken into account in the estimate of cluster sampling. From the correlations it can be concluded that the larger the value of R^2, the larger is the difference, when other parameters remain unchanged. Thus, the larger the R^2, the larger is the standard error obtained with cluster sampling (CLU), when compared to the model based estimate (SYS). Correlations with the standard error of the model were almost negligible. This may be due to the fact that standard error depends mostly on the number of sample plots, but the differences do not.

The difference between model based standard error and standard error based on quadratic form (SYS-QUAD) was also strongly correlated with the value of R^2, the correlation was –0.78 (Table 5). From the correlation it can be concluded that when the coefficient of determination increases, the difference decreases. Consequently, the standard error based on trend model is smaller than the standard error estimated with quadratic forms, when the coefficient of determination of the trend is sufficiently large. This supports the fact that the quadratic forms give overestimates of standard errors if there is large-scale correlation in the district (Matérn 1960). This situation occurs in districts 1 and 5.

The correlation of (SYS-QUAD) with parameter ν was somewhat smaller, 0.53. When the value of parameter ν increases, the difference increases, if other parameters remain unchanged. Thus, the smaller the parameter ν of the correlation function, the smaller the standard errors obtained using model based method, when compared to the quadratic form. The range of correlation, r_0, most probably affects in the same way, but since there was not much variation in the values of this parameter, this assumption could not be tested.

In districts 2, 4, and 7, the standard error estimate obtained with the trend model is even bigger than that obtained with cluster sampling. If the variation in the forests is mainly small-scale rnvariation, the quadratic form probably describes the structure of forests better than the quadratic surface model. In this case, a more flexible model, a non-parametric model for example, would probably describe the variation better than the quadratic trend.

In district 0, the smallest district, the model based estimate of standard error is smaller than that obtained with quadratic forms. This may be partly due to the trend model and the small value of parameter υ, and partly due to the fact that there are only 39 tracts in the district. For calculating the standard error accurately with quadratic forms more tracts are required. However, standard errors of the parameters of covariance function are also the bigger the less sample plots in the district.

Ordering of forest sample plots within the clusters has an effect on the differences between the general systematic sampling model and the mixed model (SYS-MIX). The correlation between the differences and the mean number of forest sample plots per cluster was –0.70, and between the differences and the mean distance of forest sample plots within the clusters –0.66. The shorter the mean distance, or the smaller the mean number of forest sample plots, the bigger the standard error of the mixed model, when compared to the general model.

The general model can take into account the differences in the ordering of forest sample plots within the cluster, but the mixed model can not. For example, if there are only two forest sample plots in the cluster, in the mixed model covariance is the same if they are subsequent to each other or at different ends of the tract.

4 Discussion

In this paper, the covariance structure in ten forestry board districts in southern Finland has been studied. The effects of the covariance structure on the mean volumes and standard errors estimated according to different methods has been considered. The results obtained show that the covariance structure in the districts, which were studied, differs from each other. The covariance structure also has an effect on the estimated results, regardless of the method used.

In the districts where large-scale variation is clear, model based estimators give smaller standard errors than quadratic forms, and in districts where the range and level of autocorrelation of random errors is large, quadratic forms give smaller estimates than model based estimators. When the trend in the population is clear, it might be reasonable to use the model based estimators, because quadratic forms are known to give conservative estimates of standard errors in this case.

Autocorrelation also had an effect on the estimated mean volumes. Model based estimators of mean volume might be preferable to classical estimators when there is variation in the number of sample plots on the tracts. The weights of the sample plots used in model based estimators are probably more precise than the weights used in classical estimators.

The problem with model based estimation is the need for computer resources. These problems can, however, be reduced by more efficient methods. Other methods, which do not need matrix inversion, have been proposed for estimating correlation functions by Vecchia (1992). These methods seem very promising because covariance functions will be needed for even larger areas in the future.

The other possibility is to use more flexible models in order to reduce the range of the correlation. If the trend model is flexible enough, autocorrelation may be reduced to within-cluster correlation. This would make calculations much easier. Besides, a more flexible model could also describe forest areas without clear large-scale trend component better than the models used in this study.

A more flexible model could be obtained by including additional variables, the distance from the sea, for example, or classifying variables in the model. The area could be stratified by a satellite image, and these strata could be used as dummy variables in the model. A non-parametric regression model for autocorrelated observations (e.g. Hart 1991) or a spatial moving averages approach (e.g. Ripley 1981) could be used in order to obtain a flexible model. In these methods, the volume of a certain point is estimated with the aid of nearby observations. Non-parametric methods might also be the most attractive solution to the calculation problems.

Model based estimates obtained were quite near the corresponding classical estimates in most districts. The quadratic trend model, however, is probably not the most reasonable choice for every district. The estimation of the model has to be made with care, and the characteristics of the area have to be considered in order to obtain a flexible enough model. The advantage of model based estimation is that the characteristics of the districts and the effects of these characteristics on estimators have to be described and studied thoroughly, and these results can be used when the results are interpreted.

References

Bellhouse, D.R. 1988. Systematic sampling. In: Krishnaiah, P.R. & Rao, C.R. (eds.). Handbook of statistics, vol. 6. Elsevier Science Publishers B.V. p. 125–145.

Cassel, C-M., Särndal, C-E. & Wretman, J.H. 1977. Foundations on inference in survey sampling. John Wiley & Sons, New York. 192 p.

Cochran, W.G. 1946. Relative accuracy of systematic and random samples for a certain class of population. Annals of Mathematical Statistics 17: 164–177.

— 1977. Sampling techniques. Third edition. John Wiley & Sons, New York. 428 p.

Cook, D.G. & Pocock, S.J. 1983. Multiple regression in geographical mortality studies, with allowance for spatially correlated errors. Biometrics 39: 361–371.

Cressie, N. 1986. Kriging nonstationary data. Journal of the American Statistical Association 81 No. 395: 625–634.

Hart, J.D. 1991. Kernel regression estimation with time series errors. Journal of the Royal Statistical Society B 53: 173–185.

Kangas, A. 1991. Updated measurement data as prior information in forest inventory. Silva Fennica 25(3): 181–191.

— 1993. Estimating the parameters of systematic cluster sampling by model based methods. Scandinavian Journal of Forest Research 8: 571–582.

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

Kvålseth, T.O. 1985. Cautionary note about R^2. American Statistician 39(4): 279–285.

Laasasenaho, J. 1982. Taper curve and volume functions for pine, spruce and birch. Seloste: Männyn, kuusen ja koivun runkokäyrä- ja tilavuusyhtälöt. Communicationes Instituti Forestalis Fenniae 108. 74 p.

Langsaeter, A. 1926–27. Om beregning av middelfeilen ved regelmessige linjetakseringer. [In Norwegian]. Medd. fra det norske skogsforsøksvesen 1(7): 5–47 and 2(8): 71–89.

Lindeberg, J.W. 1924. Über die Berechnung des Mittlerfehlers des Resultates einer Linientaxierung. Acta Forestalia Fennica 25. 22 p.

Loetsh, F., Zöhrer, F. & Haller, K. 1973. Forest Inventory. Vol. I–II. BLV Verlagsgesellschaft. 436 + 469 p.

Mandallaz, D. 1991. A unified approach to sampling theory for forest inventory based on infinite population and superpopulation models. Chair of Forest Inventory and Planning, Swiss Federal Institute of Technology (ETH), Zurich. 242 p.

Matérn, B. 1947. Metoder att upskatta noggrannheten vid linje- och provytetaxering. Summary: Methods of estimating the accuracy of line and sample plots surveys. Meddelanden från Statens Skogsforskningsinstitut, Band 36(1). 138 p.

— 1960. Spatial variation. Stochastic models and their application to some problems in forest surveys and other sampling investigations. Meddelanden från Statens Skogsforskningsinstitut, Band 49(5). 144 p.

Päivinen, R. 1987. Metsän inventoinnin suunnittelumalli. Summary: A planning model for forest inventory. Joensuu University, Publications in Sciences 11. 179 p.

Ranneby, B. 1981a. Medelfelsformler till skattningar baserade på material från den 5:e riksskogstaxeringen. Estimation of standard errors in systematic sampling. Swedish University of Agricultural sciences, Section of Forest Biometry, Report 21. 19 p.

— 1981b. Den topografiska variationen inom olika områden. En redovisning av skattade korrelationsfunktioner. Sveriges lantbruksuniversitet, Institutionen för skogstaxering, Rapport 5. 54 p.

— , Cruse, T., Hägglund, B., Jonasson, H. & Swärd,

J. 1987. Designing a new national forest survey for Sweden. Studia Forestalia Suecica 177. 29 p.
Ripley, B.D. 1981. Spatial statistics. John Wiley & Sons, New York. 252 p.
Salminen, S. 1973. Tulosten luotettavuus ja karttatulostus valtakunnan metsien V inventoinnissa. Summary: Reliability of the results from the fifth national forest inventory and a presentation of an output-mapping technique. Communicationes Instituti Forestalis Fenniae 78(6). 64 p.
— 1985. Metsien inventointimenetelmän tilastomatemaattinen perusta. Summary: The mathematical-statistical foundation of the forest inventories. Silva Fennica 19(3): 226–232.
Searle, S.R. 1971. Linear models. John Wiley & Sons, New York. 532 p.
Vecchia, A.V. 1992. A new method of prediction for spatial regression models with correlated errors. Journal of the Royal Statistical Society B 54 No. 3: 813–830.

Total of 27 references