Forest Ecosystems
a SpringerOpen Journal

Open Access

Optimal plot design in a multipurpose forest inventory

Helena M. Henttonen[1] and Annika Kangas[2*]

Abstract

Background: We explore the factors affecting the optimal plot design (size and type as well as the subsample tree selection strategies within a plot) and their relative importance in defining the optimal plot design in amultipurpose forest inventory. The factors include time used to lay out the plot and to make the tree measurements within the plot, the between-plot variation of each of the variables of interest in the area, and the measurement and model errors for the different variables.

Methods: We simulate different plot types and sizes and subsample tree selection strategies on measuredtest areas from North Lapland. The plot types used are fixed-radius, concentric and relascope plots. Weselect the optimal type and size first at plot level using a cost-plus-loss approach and then at cluster level byminimizing the weighted standard error with fixed budget.

Results: As relascope plots are very efficient at the plot level for volume and basal area, and fixed-radius plots for stems per ha, the optimal plot type strongly depends on the relative importance of these variables. The concentric plot seems to be a good compromise between these two in many cases. The subsample tree selection strategy was more important in selecting optimal plot than many other factors. In cluster level, the most important factor is the transfer time between plots.

Conclusions: While the optimal radius of plots and other parameters were sensitive to the measurement times and other cost factors, the concentric plot type was optimal in almost all studied cases. Subsample tree measurement strategies need further studies, as they were an important cost factor. However, their importance to the precision was not as clear.

Keywords: Sample, Plot, Forest inventory, Measurement, Cost, Loss

Background

Optimal inventory sampling design is a very important goal in National Forest Inventories (Mandallaz 2007). The inventory design is optimized in a sense that we wish to have the highest accuracy given a fixed budget or we wish to have the lowest cost for a given accuracy. Optimization is possible, if we make assumptions concerning the population. In an analytical setting, we need to be able to anticipate the population variance (Mandallaz & Ye 1999). It is even possible to optimize the measurements of trees in the plots, for instance to determine how many subsample trees (i.e. second-phase sample trees) to measure out of the total number of tally trees (i.e. first-phase sample trees), if we can anticipate the error in the volume estimates of the tally trees.

Defining optimal sample plot size and type analytically would require that we can anticipate the effects of the plot size and type on the population (or between-plot) variance. If the expected between-plot variation can be expressed as a function of plot size (see Freese 1961, Zeide 1980) the optimal plot size can be calculated analytically. However, such a function can only be an approximation of the between-plot variation as the relationship depends on the characteristics of the population such as spatial pattern of the trees, which cannot fully be described with a model.

In addition the expected costs, measured with time consumption as a function of plot size are needed for optimization. In fixed-radius plots the number of trees in a plot is proportional to plot area, but the time needed to check the borderline trees is proportional to the perimeter (Zeide 1980). In relascope plots, time consumption is inversely proportional to the fixed angle defined by the relascope factor. Although Kulow (1966) and

* Correspondence: Annika.kangas@luke.fi
[2]Natural Resources Institute Finland (Luke), Economics and Society unit, Yliopistokatu 6, 80100 Joensuu, Finland
Full list of author information is available at the end of the article

Grosenbaugh & Stover (1957) compared the coefficient of variation using both fixed-radius and relascope plots, they did not compare the overall efficiencies of these two types of plots related to the time spent.

While many factors affecting the accuracy can be accounted for analytically, some aspects like the spatial pattern, are more difficult. The analytical calculations usually assume a random pattern (Mandallaz 2007). Likewise, the number of subsample trees and the selection of measurements taken from each tree (e.g. height and/or upper diameter) can be difficult to account for in detail in an analytical setting. Therefore, the optimal plot size and type has most often been defined by simulating sampling in an accurately measured and mapped forest area. In the earliest studies, simulation was carried out by measuring a grid of small cells and building larger sample plots as their combination (Johnson & Hixon 1952, Mesavage & Grosenbaugh 1956). In later studies, computer simulation based on mapped data has been utilized (e.g. Kulow 1966). In a simulation based on real data, the optimal plot size is heavily dependent on the forest conditions on the area, which makes definite conclusions difficult (Mesavage & Grosenbaugh 1956).

Optimal sampling design and optimal plot design (size and type) depends highly on the purpose of an inventory. It is easy in principle to define an optimal inventory for one variable of interest such as biomass or volume with regard to measurement costs and accuracy. When the number of characteristics of interest increases, the task gets more complicated as the optimal plot number, size and type are likely to be different for each characteristic. For instance, class variables such as land use and its changes could be determined from a very small plot or even point, but volume and biomass require a larger plot. Thus, prioritizing the forest characteristics is needed if an optimal plot is to be determined.

The estimation method is also likely to have an effect: if we assume a design that is based purely on field plots, the optimal plot size and type are likely to be very different from a case where auxiliary information such as remote sensing information is used in stratification (e.g. Tomppo et al. 2014), traditional regression estimation, model-assisted estimation or model-based estimation. In these cases, the variation between plots may not be the decisive factor, but rather the correlation between the forest characteristics and the remote sensing data.

The results may also depend on the specific criterion used for defining the optimum. One option is to minimize some criterion like standard error of the estimate for a given budget constraint such as amount of time (Johnson & Hixon 1952, Mesavage & Grosenbaugh 1956). Using this approach, Johnson & Hixon (1952) concluded that while long and narrow rectangular plots tended to have smaller between-plot variation, the time

needed to lay out such plots was larger. Thus, the most efficient plots for a given amount of time were compact plots.

Another way to define the optimal plot size is to use a cost-plus-loss (CPL) approach (Hamilton 1978, Ståhl 1994). It means that the losses due to poor estimates (possibly resulting sub-optimal decisions) are calculated as a function of the uncertainty involved and these losses are added to the measurement costs described as a function of measurement time. This criterion would be ideal, if the losses due to poor estimates could be accurately defined. Often the losses are described as a function of the standard error or some other criterion (Barth & Ståhl 2012), but they could also be calculated for an actual decision problem (Eid et al. 2004). When the inventory is multipurpose, the cost-plus-loss method is more complicated (see Burkhart et al. 1978). If we were able to define the losses due to the poor estimates for each of the variables of interest (i.e. give relative weight to the errors of each variable), it is possible.

Total measurement costs can be calculated as a function of time used for each sample plot. The time depends on: 1) the time required to go to the plot and lay out the plot; 2) the total number of trees to be measured and 3) the measurements carried out for each tree. Laying out the plot means defining the plot center (or center for several sub-plots) and determining which trees belong to the plot(s). For circular or relascope plots that means checking the distance of borderline trees from the plot center with a measuring tape or an (optical) rangefinder (e.g. Loetsch et al. 1973).

The measurements needed for each tree depend on the characteristics of interest (e.g. volume, biomass, stems per ha). Typically not all characteristics needed are measured on all trees within a plot. The diameter at breast height (d1.3) is measured for all tally trees, but height, upper diameters, age, and growth are measured only for subsample trees. Thus, the measurement time also depends on the number of subsample trees within each plot, and the number of measurements carried out on each tree. As biomass and volume require additional subsample tree measurements compared to stems per ha, also the time consumption needs to be defined separately for each of the variables.
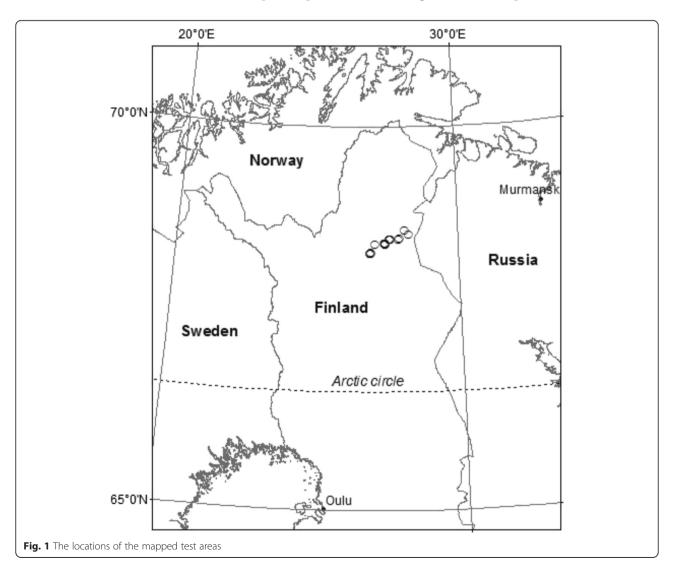
The precision of the sample plot measurements in describing the forest stand can be measured using the standard errors of the estimators of given forest characteristics, which depends, in part, on the spatial variation of the characteristics of interest within the forest. In general, the bigger the sample plot area, the larger the proportion of total variation that falls within the plot, and consequently the smaller the standard errors (e.g. Loetsch et al. 1973, Koivuniemi 2003).

Measurement and model errors for the variables used to calculate the characteristics of interest have also an effect on precision (e.g. Päivinen 1987, Ståhl et al. 2014). Their combined effect again depends on the number of subsample tree measurements and the models/methods available to generalize the subsample tree measurements to the tally trees. It may be assumed that the errors in volume / biomass for subsample trees are negligible, but not for the tally trees. It is quite possible that the model which is most efficient when all measurements are assumed error-free is not the most efficient when these errors are included (Eid 2003). Therefore, it would be best to select the models used for generalizing the subsample tree characteristics to tally trees simultaneously with deciding the number of subsample trees and the variables measured for each of them.

The aim of this study is to analyze optimal sample plot type and size with a simulation study and explore the relative effects of different factors on optimal plot

measurement strategy in the special conditions of North Lapland. The study region is partially located close to the northern timberline, where clustered spatial patterns of trees challenge the planning of an efficient forest inventory. The studied plot types were fixed-radius plots with varying radii, a combination of two concentric plots with varying radii and varying diameter limits for the larger radius, and relascope plots with varying relascope factor and maximum radii. The forest characteristics concerned were volume, basal area and stems per ha. The class variables such as forest/non-forest were excluded from the study.

Material

Measurements of 50 m x 50 m test areas were carried out in 2002 in Inari, North Finland (Figure 1). The measured areas were sampled from the plots of the 8th National Forest Inventory. In total, 18 test areas were measured, together with the planar coordinates of the

Fig. 1 The locations of the mapped test areas

trees (with d1.3 ≥ 2.5 cm) mapped with tachymeter (SOKKIA SET 4C), as well as d1.3 (in two perpendicular directions), height (h, m), and upper diameter at a height of 6 m (d6, cm) (for trees ≥ 8 m tall). An example of our data is given in Fig. 2 illustrating the spatial distribution of trees and their diameters in a mature Scots pine (*Pinus sylvestris L.*) stand with birch (*Betula pubescens*) undergrowth. Volumes for the trees were calculated using volume functions (Laasasenaho 1982). For trees with heights ≥ 8.1 m, volumes were estimated as a function of d1.3, h, and d6. When the height of a tree was less than 8.1 m, volumes were estimated as a function of d1.3 and h.

Methods

Analysis of the point patterns of trees on the 50 m x 50 m areas was carried out using the R package *spatstat* (Baddeley and Turner 2005). Our main interest was in assessing whether the point patterns could be considered random (Poisson). For this purpose, we carried out a simultaneous (simultaneous for different values of the distance r) Monte Carlo test for Ripley's K–function and L-function, which is a variance stabilizing transformation of K. Inhomogeneity was taken into account by modelling trends as a function of coordinates. In areas divided into two different stands, the stand was used as an indicator variable in modelling inhomogeneity.

Analysis of the optimal plot design was carried out at two levels: plot level and cluster level. The cluster is interpreted here as a combination of m plots, but any
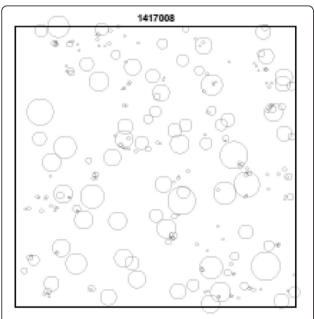

Fig. 2 An example of a mapped test area with a clustered spatial pattern. The sizes of circles are proportional to the diameters of the trees

specific spatial arrangement for the cluster is not determined. If the spatial arrangement were specified, the cluster could also be interpreted as a plot with m subplots.

In the plot level analysis, the optimal plot design was defined by minimizing the cost-plus-loss (CPL) defined for the p variables of interest for one plot. The general function to be optimized is

$$OP = \min CPL_{\{size,type,strategy\}} \\ = \min(l_1 + ... + l_P + c) \quad (1)$$

where the losses are a function of RMSE as

$$l_p = w_p RMSE_p \quad (2)$$

where w_p is the weight given to the RMSE of a given variable p. Costs c are defined as a function of the time for transfer between the plots (LT used at cluster level, assumed to be 15 min/plot), the time needed to check the borderline trees (BT assumed to be 0.5 min/tree), tally each tree in the plot (TT assumed to be 0.5 min/tree) and measure the subsample tree characteristics from each the subsample tree (ST assumed to be 4.5 min/tree) as

$$c = LT + BTn_1 + TTn_2 + STn_3 \quad (3)$$

where n_2 is the number of tally trees, n_1 is the number of borderline trees and n_3 he number of subsample trees (Päivinen 1987). A tree was defined as borderline tree if its distance from the plot center differed less than 0.5 m from the radius for a tree with a given size.

The plot level analysis was carried out so that with each plot type and size, we simulated $N = 1000$ randomly located plots (simulated plots) within each mapped 50 m x 50 m area. The accuracy of each characteristic within each area was analyzed as

$$RMSE_j = \sqrt{\sum_{i=1}^{N} \frac{\left(\hat{y}_{ji} - y_j\right)^2}{N}} \quad (4)$$

where $\hat{y}_{ji}$ is the observed forest characteristics from the area j and simulated plot i and y_j is the true value of the characteristics calculated from the whole area j. Relative RMSE was calculated by dividing the RMSE with the mean across the test areas (Table 1). The plot-level analysis $RMSE_p$ (equation 2) was the average of these RMSEs in the 18 test areas. The variables of interest were plot volume (V, m³/ha), basal area (G, m²/ha) and stems per ha (N). The simulated plots were located within the test areas so that the center point was at least 11 m from the edge so that edge corrections were not used. The possible bias resulting from this is included in the RMSE.

Table 1 Mean volumes, stem per ha and basal areas of the 50 m x 50 m mapped test areas

Testarea	V,m^3/ha	N,/ha	G,m^2/ha
1	43.4	668	8.4
2	70.6	728	13.1
3	8.2	316	1.9
4	17.0	224	3.3
5	45.0	904	9.3
6	19.0	480	4.5
7	118.8	424	17.9
8	92.4	260	12.1
9	63.9	356	11.2
10	4.0	324	1.4
11	10.6	424	2.8
12	20.3	980	5.0
13	64.6	1276	12.5
14	25.9	1388	6.5
15	44.5	1412	9.8
16	21.0	960	4.7
17	92.9	984	12.6
18	24.8	692	5.1
Mean	43.7	711.1	7.9

In the cluster level analysis, the average MSE from the plot level analysis was used as the "within-test-area" variation (Var_w). In addition, the total variation included the "between-test-areas" variation among the 18 test areas, defined as

$$Var_b = \sum_{j=1}^{18} \frac{\left(y_j - \bar{y}\right)^2}{18} \qquad (5)$$

where $\bar{y}$ is the overall mean of forest characteristic in these 18 areas (Table 1). We calculated the total variation as

$$Var_{total} = Var_b + Var_w \qquad (6)$$

where the within-area variation depends on the plot type and size but the between area variation does not.

The total variation can be used to simulate a situation where both the optimal sample plot size and the optimal number of sample plots are selected. It can thus be used to analyse if it is more useful to select a large number of small plots or a small number of large plots. In this study, we analysed the optimal plot type and size for one cluster consisted of m plots. The budget for measuring one cluster was fixed to one day work, approximately 420 min. We calculated the affordable cluster plot number m as

$$m_{case} = \frac{B}{c_{case}} \qquad (7)$$

where the c_{case} is the measurement time needed (Equation 3) for one simulated plot with a given type and size and a given subsample tree measurement strategy, and B is the total budget of measurements (in minutes per day). The relative standard error of the mean of cluster for volume (V, m^3/ha), basal area (G m^2/ha) and stems per ha (N) was calculated using the simple random sampling (SRS) formula with the number of plots m_{case} fitting the budget and the total variation Var_{total} depending on the plot type and size as

$$SE\left(\hat{\bar{y}}_{case}\right) = \sqrt{\frac{Var_{total,y}}{m_{case}}} \qquad (8)$$

Relative error was calculated by dividing with the mean across the areas (Table 1) to standardize the units. The optimal plot size, type and number of plots in a cluster was defined as the one with the minimum weighted mean of the SE%s as

$$OP = \min_{\{size,type,strategy,m_{case}\}} (w_V SE\%_V + w_N SE\%_N + w_{BA} SE\%_{BA}) \qquad (9)$$

We examined three different plot types. The first type was a fixed-radius plot with radius varying from 3 to 11 m. The second type was a combination of two concentric plots with the larger radius varying from 5 to 11 m and the smaller from 3 to 7 m. The diameter limit (DL) for trees included to the larger plots varied from 5 cm to 15 cm. The third type was a relascope plot where the relascope factor (RF) varied from 1 to 3 m^2/ha. The relascope plots were restricted to maximum radius (rmax) varying from 6 m to 11 m. Thus, the radius of the inclusion zone for trees with diameter larger than $100\, rmax \sqrt{RF} / 50$ was always equal to rmax. For the estimation of mean values per hectare using relascope plot with maximum radius see e.g. Tomppo et al. (2011). The specific plot designs tested are presented in Table 2.

We assumed that diameter on tally trees is measured in two directions. Here we tested two subsample tree selection strategies. In the first one (fixed strategy or S1), tally trees with d1.3 > 25 cm were measured as subsample trees, along with all tally trees closer than 1 m to the plot center. The assumption here is that large trees are more important subsample trees than small trees, as they contribute more to the plot volume and the variance of their volume estimates is higher than that of small trees (see discussion below). In the second strategy (relascope strategy or S2) the subsample trees were selected using a relascope factor 5 m^2/ha, also assuming that large trees are more important than small trees. We
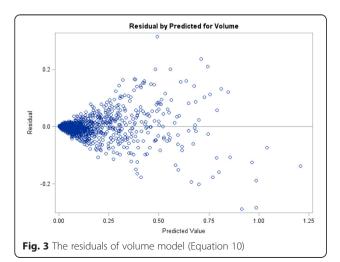
Table 2 The tested combinations of plots

Plot type	Radiusm	Radius 2m	Relascopefactor RF m^2	Diameter limit DLcm
fixed-radius	3,4,5,6,7,8,9,10 and 11	-	-	-
concentric	11	7	-	5, 7.5,10,12.5,and 15
	9	6	-	5, 7.5,10,12.5,and 15
	7	5	-	5, 7.5,10,12.5,and 15
	6	4	-	5, 7.5,10,12.5,and 15
	5	3	-	5, 7.5,10,12.5,and 15
relascope	6,7,8,9,10 and 11	-	1	-
	6,7,8,9,10 and 11	-	1.5	-
	6,7,8,9,10 and 11	-	2	-
	6,7,8,9,10 and 11	-	2.5	-
	6,7,8,9,10 and 11	-	3	-

assumed that the volume of the subsample trees could be measured error free (in fact there is error but it is assumed negligible), while for tally trees we assumed an error.

For every tree in the 18 test areas, the volume was calculated using d1.3, d6 and h as predictor variables (Laasasenaho 1982) and this was assumed to be the measured volume. Using these volumes, a simpler and less precise model

$$v = \beta_0 + \beta_1 d1.3 + \beta_2 d1.3^2 + \varepsilon \qquad (10)$$

was fitted using only d1.3 as a predictor variable. The fitted model has $R^2 = 0.9556$ and RMSE = 0.02626 m^3. The model errors were heteroscedastic (Fig. 3), which is typical for a volume model. In the analysis, the volumes of the tally trees were estimated using this simple model (10), while for the subsample trees the above mentioned measured volumes were used, to describe the effect of

Fig. 3 The residuals of volume model (Equation 10)

not measuring the height and upper diameter of each tree in the plot.

Results

The fixed-radius plots included much more measured trees than the other two plot types. In fixed-radius plot with radius 11 m, the maximum number of trees (in the 1000 replications and 18 test areas) to be measured reached 80, while in relascope plots the maximum number was below 40 and in concentric below 60 (Fig. 4). The maximum number of borderline trees was 2.39 on average for fixed-radius plots and only 0.96 for relascope plots. With fixed-radius plots, the proportion of borderline trees varied from 8.9 to 30.4 %, for relascope plots from 14.3 to 25.1 % and for concentric plots from 9.6 to 25.9 %. The variation in the number of measured subsample trees between the plot types and sizes was quite low (see also Fig. 6).

When the average relative RMSE (Equation 4) was plotted as a function of measurement times (min) for different plot types and forest characteristics, it was clear that the relascope plot type was very efficient for volume and basal area, while the fixed-radius plot was best for stems per ha (Fig. 5). The concentric sample plots seemed to be a very useful compromise, which was near optimal for all characteristics.

The strategy to select the subsample trees with an relascope factor 5 (S2) was clearly distinguishable from the strategy to select all large trees and trees closest to plot center (S1) with longer measurement times for the plot (Fig. 5). The reason for this can be seen from the number of subsample trees measured in each case with these two strategies: the relascope strategy on average produced about 0.5 more subsample trees per plot (Fig. 6). With smaller radii the difference could be as much as 1.0 subsample trees, while the differences petered out with larger plot sizes. For plots with the smallest radii
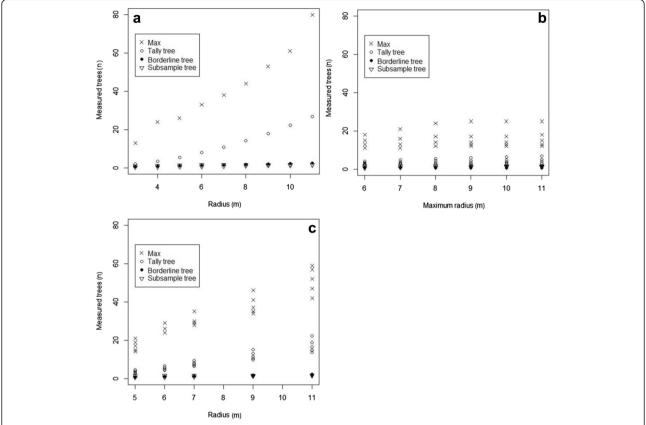
Fig. 4 The maximum number of trees in a plot, and average number of tally trees, borderline trees and subsample trees as a function of plot radius (m) with the two subsample tree selection strategies with fixed-radius (**a**), relascope (**b**) and concentric (**c**) plots. In relascope plots the variation within each relascope factor is due to varying maximum radius and in concentric plots the variation within each radius is due to varying diameter limit

the relascope strategy (S2) seemed a little more efficient with respect to the RMSE of volume while the fixed strategy (S1) was more efficient for larger radii plots (Fig. 7).

To select an optimal plot type and size for a fixed number of plots (i.e. at plot level), a cost-plus-loss analysis was carried out (Equation 1), with weight $w_P = 0.08$ for the RMSEs (Equation 2) of all three characteristics considered. For fixed-radius plots the optimal strategy of 6 m radius was very clear (CPL 20.72). The relascope strategy for selecting subsample trees was clearly less efficient than the fixed strategy (Fig. 8). For relascope plots, the differences between the subsample tree selection strategies were also very clear, as well as the differences between the relascope factors. On the other hand, the CPL did not appear to depend on the maximum radius. The optimal relascope plot had a relascope factor 1 m²/ha and a maximum radius 7 m (CPL 21.42). With concentric plots, the dependency on radius was similar but less pronounced than with fixed-radius sample plots. The optimal plot radius was a little bit larger (7 m), but

trees with d1.3 less than 15 cm were only measured within 5 m plot. This plot type produced the smallest CPL (19.81). The effect of varying the subsample tree measurement strategy was much larger than that of the diameter limit.

If the weight of RMSE for stems per ha was tripled (ceteris paribus), the optimal radius for fixed-radius plots was 7 m. In this case, the smallest CPL was obtained with fixed-radius plots (27.40). Thus, when stems per ha is important enough, the fixed-radius plot is the most efficient. For the concentric plot the optimal diameter limit changed from 15 to 5 cm. If the weight of the volume RMSE was tripled (ceteris paribus), the maximum radius of relascope plots increased to 8 m. The optimal plot was a concentric sample plot with radii 9 / 6 m and diameter limit of 15 cm (CPL = 28.31).

A more marked change occurred when the relative importance of losses compared to costs was reduced to 0.01 for all variables (Equation 2). In that case, the optimal fixed-radius plot radius was 3 m, the optimal relascope factor 3 m²/ha with a maximum radius 6 m, and
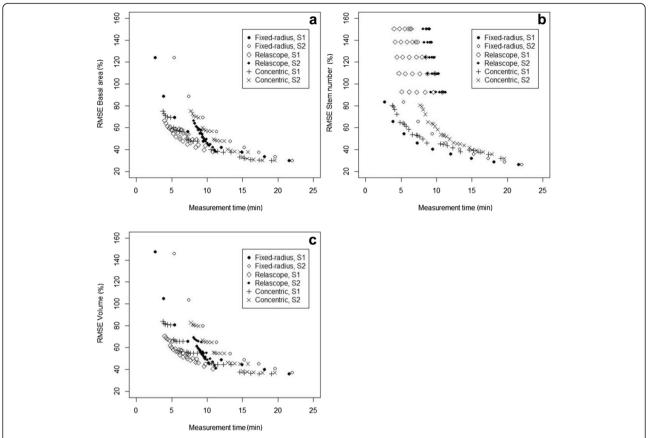
Fig. 5 The average relative RMSE as a function of measurement times (min per plot) for different plot types for basal area (**a**), stems per ha (**b**) and volume (**c**) with the two subsample tree selection strategies (S1 = fixed, S2 = relascope). In relascope plots, the vertical variation is due to varying relascope factor and the horizontal due to varying maximum radius. In fixed and concentric plots the vertical variation is due to the varying radii and the horizontal due to the diameter limit

for concentric plots the optimal radii were 5 /3 m with the diameter limit of 15 cm. In this case, the concentric plot had the smallest CPL (6.14) but the relascope plot was very close (6.18). That means that for all plot types, the optimal plot size was the smallest considered. When the weight of losses was increased compared to costs (weight 0.2 for all variables), the optimal radius for the fixed-radius plot was 8 m, the optimal relascope factor was 1 m²/ha with a maximum radius of 10 m, and the optimal concentric plot had radii of 11 / 7 m with a diameter limit 15 cm. It also had the smallest CPL (36.69).

When the budget was fixed to one day's worth of measuring minutes (420) and the sample plot number, size and type within a cluster could all be decided at the same time, the optimal combination was to measure 19 concentric sample plots with radii 7/5 m with a diameter limit of 10 cm (Fig. 9). When the time for transfer between the plots (LT) was reduced from 15 to 10 min, the optimal number of plots increased

from 19 to 25 and the diameter limit increased to 15 cm (Fig. 10). On the other hand, when LT was increased to 20 min, the optimal number of plots reduced to 14 and the optimal radii of the concentric plots increased to 9/6 m and the diameter limit increased to 12.5 cm. When the transfer time from plot to plot shortens, it is better to measure a larger number of smaller plots and vice versa.

When the measurement time of one tally tree (in Equation 3) was increased to 0.7 min and that of a subsample tree to 7 min, the optimal number of measured sample plots per cluster was reduced to 18, the optimal plot radii to 6/4 m with a diameter limit of 7.5 cm (Fig. 11). Thus, the longer it takes to measure one tree, the smaller the optimal plot size. However, the number of plots is affected less than when the transfer time is changed (Figs. 10 and 12).

The analysis of the point patterns showed that 10 out of 18 point patterns could be considered random (Poisson). However, six of these areas included parts from
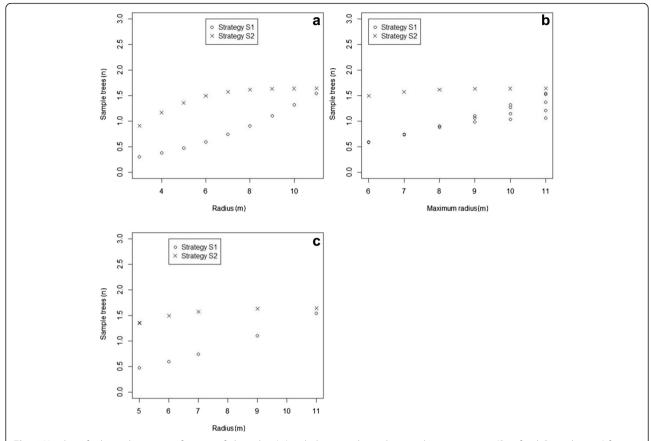
Fig. 6 Number of subsample trees as a function of plot radius (m) with the two subsample tree selection strategies (S1 = fixed, S2 = relascope) for fixed-radius (**a**), relascope (**b**) and concentric (**c**) plots. In relascope plots the variation within each maximum radius is due to varying relascope factor

more than one stand, which may have affected the spatial pattern. Seven of the point patterns were assessed as clustered. However, the clusters seemed to be quite small (< 2 m) and they were probably due mainly to birch (*Betula pubescens*) clones. Only one area showed evidence of a regular pattern. We compared the average relative RMSE's (Equation 4) of the stems per ha (Fig. 13), volume, and basal area when the test areas were classified into the different point patterns with the 6 areas divided between different stands excluded. Sampling was carried out with a fixed-radius plot with plot radius varying from 3 to 11 m. The differences between clustered and Poisson patterns were in average small, but the variation between areas was higher in clustered patterns. The relative RMSE's in the area with the regular point pattern seemed to be less sensitive to the radius of the plot.

Discussion
In this study, we analyzed the effect of plot type (fixed-radius plots, a combination of two concentric plots with a varying diameter limit, and relascope plots with varying maximum radius), different plot size (varying radii or relascope factor) and two different strategies for measuring subsample trees within plots (either all trees with $d1.3 > 25$ cm and all trees within 1 m from the plot center, or with relascope factor 5 m^2/ha). We examined three different variables, volume, basal area and stems per ha in order to reach a compromise solution that would be suitable for many other variables as well. We did not include class variables such as forest/non-forest classification or forest site or type classification, although these are important variables in forest inventory. In plot-level considerations, a very small plot or even a point would be optimal for many of these variables. Thus, including these variables to the calculations would make more sense if the whole design were optimized rather than just the plot type and size.

Relascope plots were most efficient for volume and basal area, but not as efficient for stems per ha. For stems per ha, fixed-radius plot were optimal. When the weight for stems per ha is increased enough, the fixed-radius plot becomes optimal overall. If we considered an inventory purely for stems per ha or basal
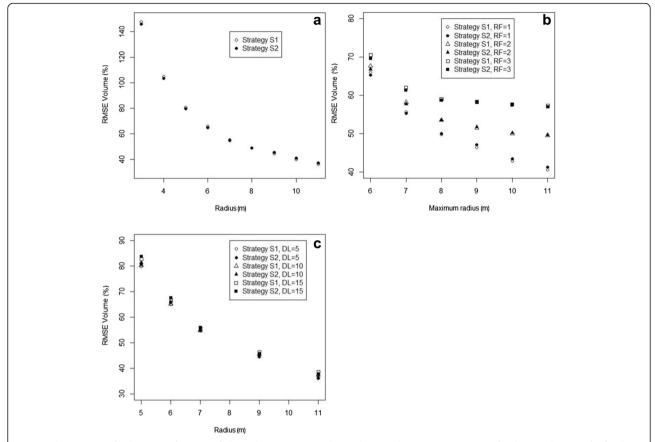
Fig. 7 Relative RMSE of volume as a function of plot radius (m) in two subsample tree selection strategies (S1 = fixed, S2 = relascope) for fixed-radius (**a**), relascope (**b**) and concentric (**c**) plots. RF = relascope factor and DL = diameter limit

area, subsample trees would not be needed at all. In fixed sized plots, measuring the diameters would neither be necessary except for borderline trees. Subsample tree selection strategies would thus be irrelevant for such an inventory. However, if the subsample tree measurement costs were removed, and the measuring cost of each tally tree would be reduced (some time would be needed to record the species and check the borderline trees), the conclusion would still be the same: relascope plot type is the best for basal area and fixed-radius plot is the best for stems per ha. While in principle there are no advantages in using fixed radius over variable radius (Stage & Rennie 1994), in relascope plots the effective plot size for small trees constituting most of the stems per ha is so small that the relascope plot type is very inefficient for stems per ha. Concentric sample plots were a good compromise between efficiency and accuracy. It also turned out that the optimal plot radius in the tested area was somewhat smaller than the one used in current Finnish NFI, 9 m.

We studied two different subsample tree selection strategies. The strategy of measuring all large trees

along with trees close to the plot center produced, on average, from 0.5 to 1.5 subsample trees in the different variations of the concentric plots, while the relascope selection strategy produced from 1.4 to 1.6 subsample trees. Although the difference may seem to be small, the strategies differed quite a lot with respect to measurement time. On the other hand, the differences in relative RMSE of volume were not large. Both the strategies acknowledge that the largest trees have largest variation in the volume estimates (Fig. 3), which makes them more attractive as subsample tree candidates. So, the relascope strategy with a larger relascope factor could have been more efficient still. The model was estimated from all the trees measured from the test area, and it therefore produced zero mean error for the whole area, but not necessarily within each diameter class. However, possible bias is implicitly accounted for in the simulations.

The results of the measurement strategy suggest that very few subsample trees would be needed. However, in this study volume was the only characteristics which required subsample tree measurements. Other variables, such as (total) biomass, might require more subsample
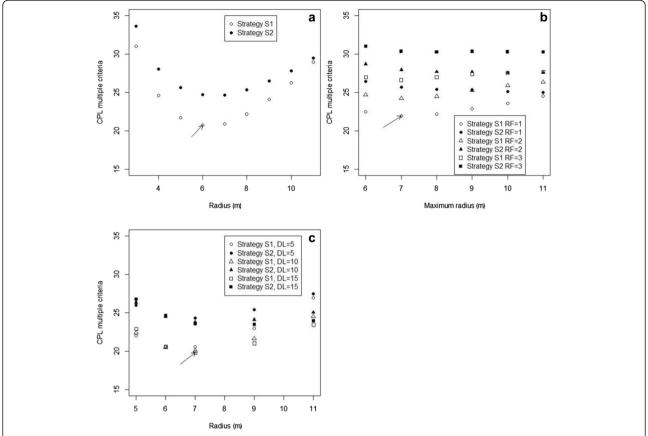
Fig. 8 Cost-plus-loss as a function of plot radius (m) with weight of 0.08 each for RMSE of G, N and V in the loss function in two subsample tree selection strategies (S1 = fixed, S2 = relascope) for fixed-radius (**a**), relascope (**b**) and concentric (**c**) plots. The optimal plot size (minimum CPL assuming fixed plot number) with each plot type is marked with an arrow
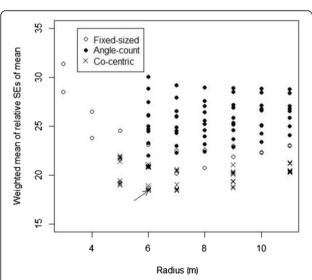

Fig. 9 The weighted mean of relative standard errors for mean volume, basal area and stems per ha as a function of plot radius (maximum radius for relascope plots). The number of plots is based on a budget of 420 min for one cluster. Assumed costs (Equation 3) are LT = 15, BT = 0.5, TT = 0.5 and ST = 4.5 min. The optimal number of plots for a cluster using the optimal plot type and size (marked with an arrow) is 19

trees as biomass models using only d1.3 as predictor are generally less precise than similar volume models. We also considered only temporary plots here. If we had analysed growth of the trees using permanent plots, more subsample tree measurements might prove to be needed, as the estimated growth per tree would be more reliable. These issues remain to be studied in the future.

The optimal plot size and number is quite sensitive to the assumed times to move from plot to plot or of measuring the trees. The concentric plot type was the best plot type for both plot-level and cluster-level calculations, and practically irrespective of the changes in the parameters in the cost function or the weights of different variables. On the other hand, this result can depend on the conditions in Lapland, and in other condition such as southern Finland or tropical areas some other plot type would be optimal.

We did not consider the effect of diameter distribution in this study, but it may also have an effect on the optimal plot type and size. In our northern data, 45 cm was the largest diameter at breast height (the maximum diameter within one test area varied from 19 to
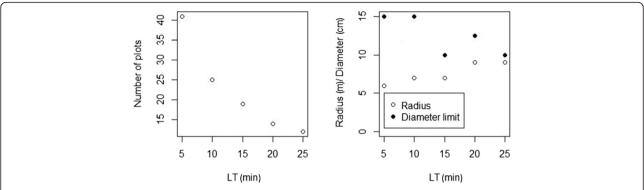
Fig. 10 Sensitivity of the optimal number of plots, diameter limit and plot radii for concentric plot as a function of time for transfer between the plots (LT, min)

45 cm). If the variation had been greater, the variation among the smaller sized plots would most likely have been higher. This also remains to be studied in the future.

We did a preliminary analysis about the effect of point pattern on accuracy. The areas with clustered patterns seemed to have higher between-stands variation in RMSE, although the size of the tree clusters was quite small compared with the tested plot radii. The RMSE on one study area with regular point pattern seemed to be less sensitive to the plot radii than the Poisson and clustered patterns. This might be of importance in the planning of inventories in the
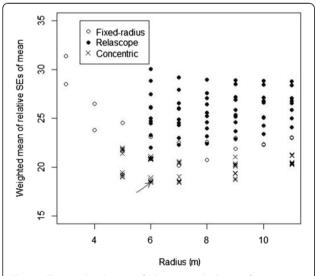

Fig. 11 The weighted mean of relative standard errors for mean volume, basal area and stems per ha as a function of plot radius (maximum radius for relascope plots). The number of plots is based on a budget of 420 min for one cluster. Assumed costs (Equation 3) are LT = 15, BT = 0.7, TT = 0.7 and ST = 7.0 min. The optimal number of plots for a cluster using the optimal plot type and size (marked with an arrow) is 18

future, since the area and volumes of regular planted forests is rapidly increasing in Finland. The effect of point patterns on optimal sampling needs further study and modelling efforts.

If remote sensing material was used as auxiliary data and a model-assisted or model-based framework was employed instead of e.g. simple random sampling, a larger plot size might be optimal (see e.g. Hofstadt et al. 2015). This is because we assume the correlation between the remote sensing data and plot data to be higher with larger plots due to e.g. co-registration errors. Moreover, remote sensing registers crowns rather than stems, and crowns of trees included into the plot will often be partly outside the plot boundaries, and respectively the crowns of trees not included into the plot will be partly inside the plot boundaries. Within larger plots, the effect of crown overlapping should be smaller. This also remains to be studied in the future.

In this study, we searched for an optimal plot type and size at plot level, i.e. for the case when the number of plots is fixed, and for a case where the optimal number of plots in a cluster was defined simultaneously with the plot type and size, i.e. for the case where the number of clusters is fixed. The analysis is valid for a wide range of sampling designs used. However, the resulting optimal plot design could be sub-optimal if also the sampling design and total plot number were simultaneously optimized. For instance, it might be better to measure a large number of small clusters (like half day clusters) rather than a small number of large clusters. Or it might be better to measure less and larger plots if remote sensing material were used as auxiliary data. Unfortunately the data we had available is not large enough for such analysis.

The relative importance of the optimal plot type and size in defining the optimal sampling design has not been defined. Based on our results, we would recommend that
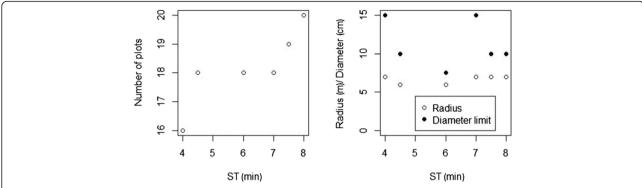
Fig. 12 Sensitivity of the optimal number of plots, diameter limit and plot radii for concentric plot as a function of time needed to measure one subsample tree (ST, min)

the whole chain of decisions from measurements, plot type, plot size, number of plots (total and/or within a cluster), number of clusters, cluster design (spatial arrangement of the plots within a cluster), sampling design and the estimation method should be simultaneously defined. Such analysis would, however, require a very large area that is measured in detail, with very large costs. Nowadays, a simulated forest might be a better option (e.g. Päivinen 1987). The design has often been optimized using a forest map based on a satellite image (e.g. Tomppo et al. 2010, 2011), but while that approach allows for selecting the optimal cluster design and number of plots within a cluster, it does not include enough information for selecting the optimal plot type. A mapped forest area based on individual tree detection from a lidar data (Holopainen et al. 2013) might provide a good starting point for a data where total optimization is possible.

Conclusions

While the optimal radius of a plot and other design parameters were quite sensitive to the measurement time and other cost factors, the concentric plot type was optimal in almost all studied cases. It is important to select a plot size that would be near optimal in many different conditions. Here, for instance, a 6–7 m radius and 10 cm diameter limit was optimal or near optimal option in most calculations. Yet, it needs to be noted that the results were calculated for Northern Finland, and elsewhere a separate optimality analysis would be needed.

The more weight is given to the costs compared to the RMSEs of the variables of interest, the smaller the optimal plots with a fixed plot number. With fixed budget, having more, smaller plots is optimal, if the transfer time between the plots is short. However, the distance between the plots within a cluster and therefore also the transfer time needs to be selected long enough to avoid high autocorrelation between the plots.

Subsample tree selection and measurement strategies need further studies, as subsample trees are a quite important cost factor but their importance to the accuracy of the final results was not as clear. The errors for tally trees had little impact on the accuracy of volume, but when other variables such as volume growth is analyzed, the subsample tree measurements may be of greater importance.
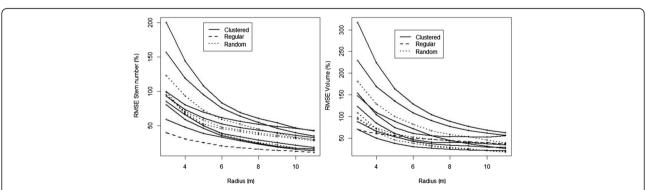

Fig. 13 RMSE (%) of stems per ha and mean volume for test areas with different point patterns as a function of plot radius (m) for fixed-radius plot. Test areas completely within one stand included

Authors' contributions
HH was responsible for designing the experiment and collecting the data, analyzing the point patterns, and main responsible for coding the simulator for sample plots. AK was main responsible for calculating and analyzing the results and writing the article. Both authors read and approved the final manuscript.

Author details
[1]Natural Resources Institute Finland (Luke), Economics and Society unit, Jokiniemenkuja 1, 01370 Vantaa, Finland. [2]Natural Resources Institute Finland (Luke), Economics and Society unit, Yliopistokatu 6, 80100 Joensuu, Finland.

References
Baddeley A, Turner R (2005) Spatstat: an R package for analyzing spatial point patterns. J Stat Software 12(6):1–42. ISSN: 1548–7660. URL: http://www.jstatsoft.org/article/view/v012i06

Barth A, Ståhl G (2012) Determining sample size in national forest inventories by cost-plus-loss analysis: an exploratory case study. Eur J For Res 131:339–346

Burkhart HE, Stuck RD, Leuschner WA, Reynolds MA (1978) Allocating inventory resources for multiple-use planning. Can J Forest Res 8:100–110

Eid T (2003). Model validation by means of cost-plus-loss analyses. In: Amaro A, Reed D, Soares P (eds) Modelling forest systems. Cambridge USA: CABI Publishing. pp 295–305.

Eid T, Gobakken T, Næsset E (2004) Comparing stand inventories for large areas based on photo-interpretation and laser scanning by means of cost-plus-loss analyses. Scand J For Res 19:512–523

Freese F (1961) Relation of plot size to variability: an approximation. J For 59:679

Grosenbaugh LR, Stover WS (1957) Point-sampling compared with plot-sampling in Southeast Texas. For Sci 3:2–14

Hamilton DA (1978) Specifying precision in natural resource inventories. In: Integrated inventories of renewable resources: proceedings of the workshop. Tuscon, Arizona Usa: USDA Forest Service, General technical report RM-55: 276–281.

Hofstadt EH, Gobakken T, Solberg S, Kangas A, Ene L, Mauya E, Næsset E (2015) Relative efficiency of ALS and InSAR for biomass estimation in Tanzanian rainforest. Remote Sens 7:9865–9885

Holopainen M, Kankare Vi, Vastaranta M, Liang, X, Lin Y, Vaaja M, Yu X, Hyyppä J, Hyyppä H, Kukko A, Tanhuanpää T, Alho P (2013) Tree mapping using airborne, terrestrial and mobile laser scanning – A case study in a heterogeneous urban forest. Urban Forestry & Urban Greening 12:546–553. doi:10.1016/j.ufug.2013.06.002 DOI:10.1016/j.ufug.2013.06.002#doilink

Johnson FA, Hixon HJ (1952) The most efficient size and shape of plot to use for cruising in old-growth douglas-fir timber. J For 1:17–20

Koivuniemi, J (2003) Metsiköihin ja paikannettuihin koealoihin perustuvan kuvioittaisen arvioinnin tarkkuus. Summary: The accuracy of the compartmentwise forest inventory based on stands and located sample plots. Doctoral thesis. Publications of the Department of Forest Resource Management 36. University of Helsinki, Helsinki, p 160

Kulow DL (1966) Comparison of forest sampling designs. Journal of Forestry July 469–474.

Laasasenaho J (1982) Taper curve and volume functions for pine, spruce and birch. Commun. Inst. For. Fenn. 108. p 72

Loetsch F, Zöhrer F, Haller KE (1973) Forest Inventory Volume 2. BLV Verlagsgesellschaft, München, p 469

Mandallaz D, Ye T (1999) Forest inventory with optimal two-phase, two-stage sampling schemes based on the anticipated variance. Can J Forest Res 29:1691–1708

Mandallaz D (2007) Sampling techniques for forest inventories. Chapman & Hall. p 256

Mesavage C, Grosenbaugh LR (1956) Efficiency of several cruising designs on small tracts in North Arkansas. Journal of Forestry September 569–576.

Päivinen R (1987) Metsän inventoinnin suunnittelumalli. [A planning model for forest inventory, In Finnish]. University of Joensuu publications in Sciences, Joensuu, N:o 11, p 179

Stage AR, Rennie JC (1994) Fixed-radius plots vs. variable-radius plots. J For 92:20–24

Ståhl G (1994) Optimizing the utility of forest inventory activities. Ph.D. thesis, Swedish University of Agricultural Sciences, Department of Biometry and Forest Management, Umeå.

Ståhl G, Heikkinen J, Petersson H, Repola J, Holm S (2014) Sample-based estimation of greenhouse gas emissions from forests - a new approach to account for both sampling and model errors. For Sci 60(1):3–13

Tomppo E, Gschwanter T, McRoberts RE, Lawrence M, Editors (2010). National forest inventories – pathways for common reporting. Springer. ISBN 978-90-481-3232-4.

Tomppo E, Heikkinen J, Henttonen HM, Ihalainen A, Katila M, Mäkelä H, Tuomainen T, Vainikainen N (2011). Designing and conducting a forest inventory - case: 9th National Forest Inventory of Finland. Springer, Managing Forest Ecosystems 21, p 270. ISBN 978-94-007-1651-3

Tomppo E, Malimbwi R, Katila M, Mäkisara K, Henttonen HM, Chamuya N, Zahabu E, Otieno J (2014) A sampling design for a large scale forest inventory: case Tanzania. Can J Forest Res 44:931–948

Zeide B (1980) Plot Size Optimization. For Sci 26:251–257