RESEARCH   Open Access

The effect of spatial structure of forests on the precision and costs of plot-level forest resource estimation

Henrike Häbel[1], Mikko Kuronen[1], Helena M. Henttonen[1], Annika Kangas[2] and Mari Myllymäki[1]*

Abstract

Background: We investigated how the precision and costs of forest resource estimates for sample plots of different type and size depend on the spatial structure of forests and jointly studied the effects of tree density and size distribution. Statistically thinking, the trees in a forest can be regarded as a point pattern. Based on the spatial properties of the point pattern, we classified the forests into clustered, random, and regular. We used empirical data from 396 mapped forest plots from Finland. The variance of the unbiased Horvitz-Thompson estimator and expected costs of the basal area and tree density estimation were calculated for 99 different sample plots of different type and size in each of the 396 forest plots. Further, we considered the estimation of the change between two time points for a subset of the data.

Results: The precision and expected cost depended on the tree size distribution and spatial pattern of trees. While large sample plots are advisable for clustered forests or the monitoring of young forests with small trees, we see potential for measuring smaller sample plots in regular forests. The choice of sample plot was more important in clustered forests, where also the variability of the expected costs was higher.

Conclusions: If the spatial structure of forests could be predicted accurately and precisely prior to field measurements, for instance from remote sensing data, the precision of forest inventories could potentially be improved or costs decreased by allowing the sample plot size and type to vary from one forest stand to another. When using a compromise sample plot over a large region and a long inventory rotation, optimizing the sample plot for one time point ignores possible changes in forest structures caused by changes in forest management practices.

Keywords: Concentric circular plot, Expected costs, Fixed radius plot, Forest inventory, Horvitz–Thompson estimator, Relascope plot, Sample plot, Spatial pattern of trees, Spatial structure

Background

Typically, a sampling-based national forest inventory (NFI) is carried out to provide statistics, for example for national or regional forest programs, sustainability assessments, investment calculations for forest industry, and reporting to international conventions (Tomppo et al. 2010). Forest management inventories (FMI), on the other hand, are carried out to support forest owners in their strategic planning as well as aid decision making concerning harvests and silvicultural measures.

NFIs are based on a set of field plots, for instance in Finland about 15 000 plots are measured each year (Kangas et al. 2018) totaling to 75 000 plots during each five year inventory rotation. In recent years the forest management inventories have been carried out using airborne laser scanning (ALS) data and field plots (e.g. Næsset (2004)). The field plots used may be the NFI field plots or plots measured as a separate measurement campaign. Thus, the field plot measurement is a large investment, and it is important to make it as cost-efficiently as possible.

In NFI, the inventory design is optimized in the sense that we wish to have the highest accuracy and precision given a fixed budget or we wish to have the lowest cost

*Correspondence: mari.myllymaki@luke.fi
[1]Natural Resources Institute Finland (Luke), Latokartanonkaari 9, FI-00790 Helsinki, Finland
Full list of author information is available at the end of the article

for a given precision (Päivinen 1987). In a cluster design, that entails selecting the number of clusters or distance between them, the number of plots in each cluster and the distance between them, and the type and size of plots within each cluster. Optimization is possible, if we make assumptions concerning the population (Mandallaz 2007). In an analytic setting, we need to be able to anticipate the between-plot variance (Mandallaz and Ye 1999).

Defining optimal sample plot size and type analytically would require that we can anticipate the effects of the plot size and type on the between-plot variance. If the expected between-plot variation could be expressed as a function of plot size (see Zeide (1980)) the optimal plot size could be calculated analytically. The optimal size, however, also depends on the size distributions of the forest stands as well as on the spatial patterns of trees.

While some factors affecting the accuracy and precision can be accounted for analytically, other aspects like the spatial structure of forests are more complicated. The analytic calculations usually assume a random pattern (Mandallaz 2007). In Finland, however, regular patterns of trees have been observed more commonly (57%) than random (25%) and clustered (18%) patterns (Tomppo 1986). Therefore, the optimal plot is typically addressed using simulation (e.g. Henttonen and Kangas (2015)).

In this study based on field data from different types of forests, we compared the precision of the basal area and tree density estimation and costs for various sample plots of different type and size. We used the unbiased Horvitz-Thompson estimator to estimate the basal area and tree density and obtained its variance by integrating over the forest plot. The resulting standard deviation and corresponding costs of various sample plots were investigated with respect to the spatial structure of forests and other stand variables such as tree density, basal area and diameter distribution that are potentially correlated with the spatial structure (see e.g. Tomppo (1986) p. 40 ff.). By classifying our data, we studied how the spatial structure of forests and other stand variables affect precision and costs of different sample plots. Further, we examined the effect of small trees on the spatial structure, precision and costs.

Methods
Materials
We used data of tree locations and diameters measured at 1.3 m (*dbh*) from different types of forest stands, including experiments of uneven-aged forest management and thinnings from below resulting in different spatial patterns of trees. Also data from natural forests and peatlands were included. The total number of forest plots from different data sources was 396 (Fig. 1). In order to distinguish this data from the sample plots, they are henceforth referred to as forest plots.
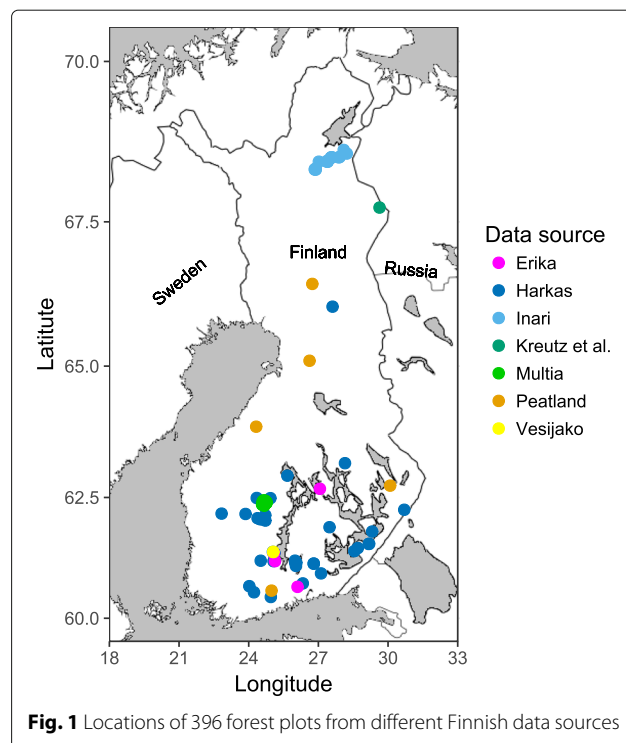

Fig. 1 Locations of 396 forest plots from different Finnish data sources

Erika
The Erika data set consists of 21 experimental plots of 40 m × 40 m in managed, uneven-aged Norway spruce dominated forests at five locations in southern Finland (Eerikäinen et al. 2014). Between 1991 and 2012, the plots were measured usually four to five times with five years' intervals, recording the locations, tree species, and size characteristics (*dbh* and height) of all trees ≥ 10 cm tall. Trees with *dbh* ≥ 0.1 cm were included in our calculations. Only the fourth measurement of each plot was included in our main study. These measurements were taken either in 2007 or 2012. From one plot with only two measurements, the data from 1992 was used.

Harkas
The Harkas data set comprises 312 plots from 39 long-term thinning experiments mainly in southern and central Finland (Mäkinen and Isomäki 2004). These experiments have been established in even-aged, pure, or almost pure Scots pine or Norway spruce stands growing on mineral soil. The recording of tree locations on plots started in the late 1980s. We used the locations and tree *dbh*s from the most recent measurement for each plot, since the more recent locations were measured with more accurate methods, i.e. tachymeter, instead of distance (measuring tape) and compass direction used earlier. Only plots where both tree locations and plot corners had been mapped were included. Plot areas were estimated using the coordinates of plot corners and therefore these areas are not

exactly equal to those presented in earlier reports. The median area of the plots was 1004 m^2 (range 788 m^2 - 2460 m^2). The minimum *dbh* in the Harkas measurements was 4.5 cm.

Inari

The Inari data set consists of 18 forest plots of size 50 m × 50 m measured on the locations of plots from the 8th National Forest Inventory in northernmost Finland (Henttonen and Kangas 2015). The coordinates (tachymeter measurement) and *dbh* of all trees with *dbh* ≥ 2.5 cm were recorded.

Multia

The Multia data set consists of 30 field plots of size 32 m × 32 m measured in 2014 in Multia and Keuruu, southern Finland (Tomppo et al. 2016). These data were collected for validation of a remote-sensing based forest resource estimation. Therefore, plots were subjectively located to young thinning stands, advanced thinning stands and mature stands. The coordinates and *dbh*s of trees with *dbh* ≥ 2.5 cm were measured using Sonar caliper.

Various additional data sets

The Various additional datasets consist of large plots measured in different types of forests. Data from Vesijako, southern Finland, comprises two plots of size 100 m × 100 m and 70 m × 130 m. The plots were located in mixed coniferous advanced thinning stands dominated by Norway spruce and Scots pine, respectively. The coordinates (tachymeter measurement) and size characteristics (*dbh* and height) of all trees were recorded. The smallest observed *dbh* in these plots was 5.7 cm.

The data from Kreutz et al. (2015) was collected from Värriö Strict Nature Reserve in northeastern Finland and it consists of three plots of size 300 m × 40 m located in natural forests at different successional stages. In measuring tree *dbh*s they were rounded to the closest cm. The minimum *dbh* was 1 cm.

Hökkä et al. (2008) collected mapped tree data from seven Scots pine dominated stands growing on drained peatlands in different parts of Finland. We used their data from a total of 10 plots in five stands. Similarly to the processing of Harkas data set, we left out some parts of the data because of missing or unclear plot corner or tachymeter location measurements and the plot areas were thus not equal to the report of Hökkä et al. (2008). The median area of plots in our calculations was 6392 m^2 (range 1604 m^2 - 15620 m^2). The minimum *dbh* was 4.5 cm.

Sample plot types and sizes

We considered 99 different sample plots, including different specifications for the circular fixed radius determining the overall size, relascope and circular concentric plots specifying the type as follows:

(i) Fixed radius plots with the radius $r_{max} = 3, 4, \ldots, 11$ m in which all trees above a certain *dbh* are measured (9 different plots).

(ii) Relascope plots with the maximum radius $r_{max} = 3, 4, \ldots, 11$ m and basal area factors $q = 1, 1.5, 2, 2.5, 3$ (45 different plots). A tree is included in the sample plot if its distance to the sample plot center is at most $\min(r_{max}, 50\, dbh/\sqrt{q})$ (see Tomppo et al. (2011), p. 25 f.) for more details).

(iii) Concentric plots with radii $(r_1, r_{max}) = (1, 3), (2, 4),$ $(3, 5), (4, 6), (5, 7), (5, 8), (6, 9), (6, 10), (7, 11)$ m for the two circles where the trees with the *dbh* smaller than 5, 7.5, 10, 12.5, 15 cm are measured only from the smaller circle with radius r_1 (45 different plots).

Precision of forest resource estimation

We assumed that the sample plot location *s* is chosen uniformly randomly from the observed forest plot window $W \subset \mathbb{R}^2$ and that the forest characteristic Y is estimated using the unbiased Horvitz–Thompson estimator (Horvitz and Thompson 1952)

$$\hat{Y}(s) = \frac{1}{|W|} \sum_{i=1}^{n} \frac{I_i(s) Y_i}{\pi_i}, \tag{1}$$

where n is the finite number of observed trees in the observation window W of the forest plot, $I_i(s)$ is the indicator that tree i is in the sample plot centered at s, Y_i is the characteristic of tree i, π_i is the inclusion probability of tree i, and $|W|$ denotes the forest plot area. We considered the estimation of the tree density measured in number of stems per hectare ($Y_i = 1$) and the basal area per hectare $\left(Y_i = \pi(dbh_i/2)^2\right)$.

We measured the precision of the forest resource estimation by the variance of the estimator (1). Based on Theorem 4.2.1. from Mandallaz (2007), the variance is

$$V[\hat{Y}(s)] = \frac{1}{|W|^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{Y_i Y_j (\pi_{ij} - \pi_i \pi_j)}{\pi_i \pi_j}, \tag{2}$$

where π_{ij} is the inclusion probability for the pair of trees i and j, given by

$$\pi_{ij} = \mathbf{E}[I_i(s) I_j(s)] = \frac{|B(x_i, r_i) \cap B(x_j, r_j) \cap W|}{|W|}, \tag{3}$$

where x_i is the location of tree i, r_i is the inclusion radius for tree i given by the sample plot type and size, and $B(x_i, r_i)$ is the inclusion zone of tree i, i.e. a circle with centre at x_i and radius r_i. For $i = j, \pi_{ii} = \pi_i = |B(x_i, r_i) \cap W|/|W|$. The intersection areas in (3) were computed approximately using the R library spatstat (Baddeley et al. 2015). Consequently, the variance (2) was calculated for each sample plot in each forest plot described above going through all pairs of trees in a forest plot.

We further considered the estimation of the change from a time point 1 to time point 2 for the Erika data with measurements repeatedly over time, and estimated the change simply as the difference of the estimates at the two time points, $\hat{Y}^{(2)}(s) - \hat{Y}^{(1)}(s)$. The formula for the variance of the difference is

$$
V[\hat{Y}^{(2)}(s) - \hat{Y}^{(1)}(s)] = V\left[\hat{Y}^{(2)}(s)\right]
$$
$$
+ V\left[\hat{Y}^{(1)}(s)\right] - 2\mathrm{Cov}\left[\hat{Y}^{(1)}(s), \hat{Y}^{(2)}(s)\right],
$$

(4)

where

$$
\mathrm{Cov}\left[\hat{Y}^{(1)}(s), \hat{Y}^{(2)}(s)\right]
$$
$$
= \frac{1}{|W|^2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \frac{Y_i^{(1)} Y_j^{(2)} (\pi_{1i2j} - \pi_{1i}\pi_{2j})}{\pi_{1i}\pi_{2j}},
$$

(5)

where 1i and 2j refer to tree i at time point 1 and tree j at time point 2, respectively.

Costs for the comparison of sample plot types and sizes
We assumed that the cost (time in minutes) to measure tree i in the sample plot of certain type and size was

$$
c_i = 0.5 \cdot \mathbf{1}(s \in B(x_i, r_i) \cap W)
$$
$$
+ 0.5 \cdot \mathbf{1}\left(s \in [B(x_i, 1.1r_i) \setminus B(x_i, 0.9r_i)] \cap W\right),
$$

(6)

where s is the sampling location and r_i is the inclusion radius of tree i. Thus, we assumed simply a fixed cost per measured tree (0.5 min) and an additional cost per borderline tree (0.5 min). A tree was defined to be a borderline tree if its distance from the sample location was between $0.9r_i$ and $1.1r_i$. Given the single tree cost (6), the expected cost for measuring the sample plot is

$$
\mathbf{E} \sum_{i=1}^{n} c_i = \sum_{i=1}^{n} \mathbf{E}c_i,
$$

(7)

where

$$
\mathbf{E}c_i = 0.5 \frac{|B(x_i, r_i) \cap W|}{|W|}
$$
$$
+ 0.5 \frac{|[B(x_i, 1.1r_i) \setminus B(x_i, 0.9r_i)] \cap W|}{|W|}.
$$

Similarly as the variance (2), the expected cost (7) was calculated for each sample plot in each forest plot.

The expected cost for trees of the sample plot in a size class $D_k = [d_k, d_{k+1})$, $k = 1, \ldots, K$, is

$$
\mathbf{E} \sum_{i=1}^{n} c_i \mathbf{1}(dbh_i \in D_k) = \sum_{i=1}^{n} \mathbf{1}(dbh_i \in D_k) \mathbf{E}c_i.
$$

Classification of forest structure
In order to study the effect of spatial structure of forests on the precision and costs of different sample plots, the forest plots were divided into different groups of similar stand variables.

For the classification of the spatial structure of a forest plot, tree locations (at stem center) are mathematically expressed as point patterns $(x_1, \ldots, x_n)$ with a finite number of n trees observed on a forest plot window W. Each point pattern is analyzed as a realization of a point process X, which is assumed to be translation and rotation invariant. In what follows, the terms point and tree location can be used interchangeably. Likewise, the intensity λ of the point process is equal to the tree density, albeit per m^2.

The tree locations were divided into clustered, random and regular patterns by utilizing the *L*-function, which is the variance stabilizing transformation of Ripley's *K*-function (Chiu et al. 2013, Chapter 4.6) given by

$$
L(r) = \sqrt{\frac{K(r)}{\pi}} \quad \forall r \geq 0.
$$

(8)

The function $\lambda K(r)$ gives the expected number of points of X within a circle $B(o, r)$ around a typical point o and radius r (in m) without counting o itself given that there is a point of X in o.

In the completely spatially random (CSR) case with no interaction between the points, $L(r) - r = 0$ for all $r \geq 0$. This fact can be used in a test for CSR based on the test statistic

$$
\tau = \max_{r \leq r_t} |\widehat{L}(r) - r|
$$

(9)

with Ripley's isotropic edge corrected estimator $\widehat{L}$. The CSR hypothesis can be rejected at a 5% significance level if

$$
\tau > \frac{1.45\sqrt{|W|}}{n}
$$

(10)

(Chiu et al. 2013, pp. 57 f., 139 ff.). Due to size limitations of some forest plots in the Harkas data, distances up to $r_t = 5$ m were taken into account for a short range classification.

For the analysis, the statistical software R version 3.4.4 (Core Team 2018) was used together with the package `spatstat` (Baddeley et al. 2015).

Results
Relationship between spatial structure and stand variables
In order to understand the relationship between the spatial structure of forests and other stand variables in different types of forests, we calculated the average (standard deviation) tree density, mean basal area weighted *dbh*, and basal area of the forest plots and coupled them to the results of the spatial structure classification (Table 1). The Erika data with the smallest minimum *dbh* of 0.1 cm have mainly clustered plots with a relatively high tree density

Table 1 Average tree density (standard deviation) in trees/ha, average basal area weighted mean *dbh* in cm (standard deviation), average basal area (standard deviation) in m^2/ha, and number of clustered, random, and regular plots for the Erika (21 plots), Harkas (312 plots), Inari (18 plots), Multia (30 plots), and various other data sources (15 plots)

	Tree density	Min *dbh*	Mean *dbh*$_{BA}$	Basal area	Structure		
					Clustered	Random	Regular
Erika	1571 (630)	0.1	26 (3)	23 (4)	15	4	2
Harkas	949 (536)	4.5	23 (4)	30 (9)	2	2	308
Inari	711 (392)	2.5	18 (6)	8 (5)	12	4	2
Multia	2855 (1866)	2.5	15 (6)	23 (8)	10	2	18
Various	1000 (386)	0.5-5.7	16 (7)	17 (8)	6	0	9

and mean *dbh*. The Harkas data including the thinning experiments have regular tree patterns, a lower average tree density, and the largest basal area. The Inari data with a minimum *dbh* of 2.5 cm have relatively few trees and appear mostly clustered, whereas the Multia data with the same minimum *dbh*, but higher tree density, have both clustered and regular patterns. In contrast to Tomppo (1986), where smaller forest plots where considered, only 12 forest plots (3%) were classified as random. This small proportion of random forests results from large forest plots of this study that facilitate the detection of clustering and regularity due to more observations.

Given the classification according to spatial structure, the forest plots were partitioned according to standard deviation of *dbh*, basal area and tree density using the recursive partitioning function implemented in the `rpart` R package (Therneau et al. 2017). We concluded from the partitioning for forest structure that clustered patterns had a larger *dbh* range than regular patterns, whereas regular patterns tended to have larger trees with about the same size. Consequently, a grouping with respect to *dbh* standard deviation was not informative for this study since one group would mainly contain clustered and the other mainly regular forests. In our data, all forest plots with a tree density $\geq$ 3214 trees/ha were classified as clustered.

Based on the partitioning, each forest plot was classified as a plot with small ($<$ 9.8 m^2/ha), medium (in [9.8,32) m^2/ha) or large ($\geq$ 32 m^2/ha) basal area. We further defined different development classes also taking Tomppo et al. (2011, Table 2.17) into consideration. The resulting development classes are based on the mean basal area weighted *dbh* (*dbh*$_{BA}$) and tree density:

- *seedling stand*: mean *dbh*$_{BA}$ $<$ 8 cm or density $>$ 3214 trees/ha,
- *young stand*: mean *dbh*$_{BA}$ in [8,26] cm and density in (1500,3214] trees/ha,
- *advanced stand*: mean *dbh*$_{BA}$ in [8,26] cm and density in (500, 1500] trees/ha,
- *mature stand*: mean *dbh*$_{BA}$ $>$ 26 cm or tree density $\leq$ 500 trees/ha.
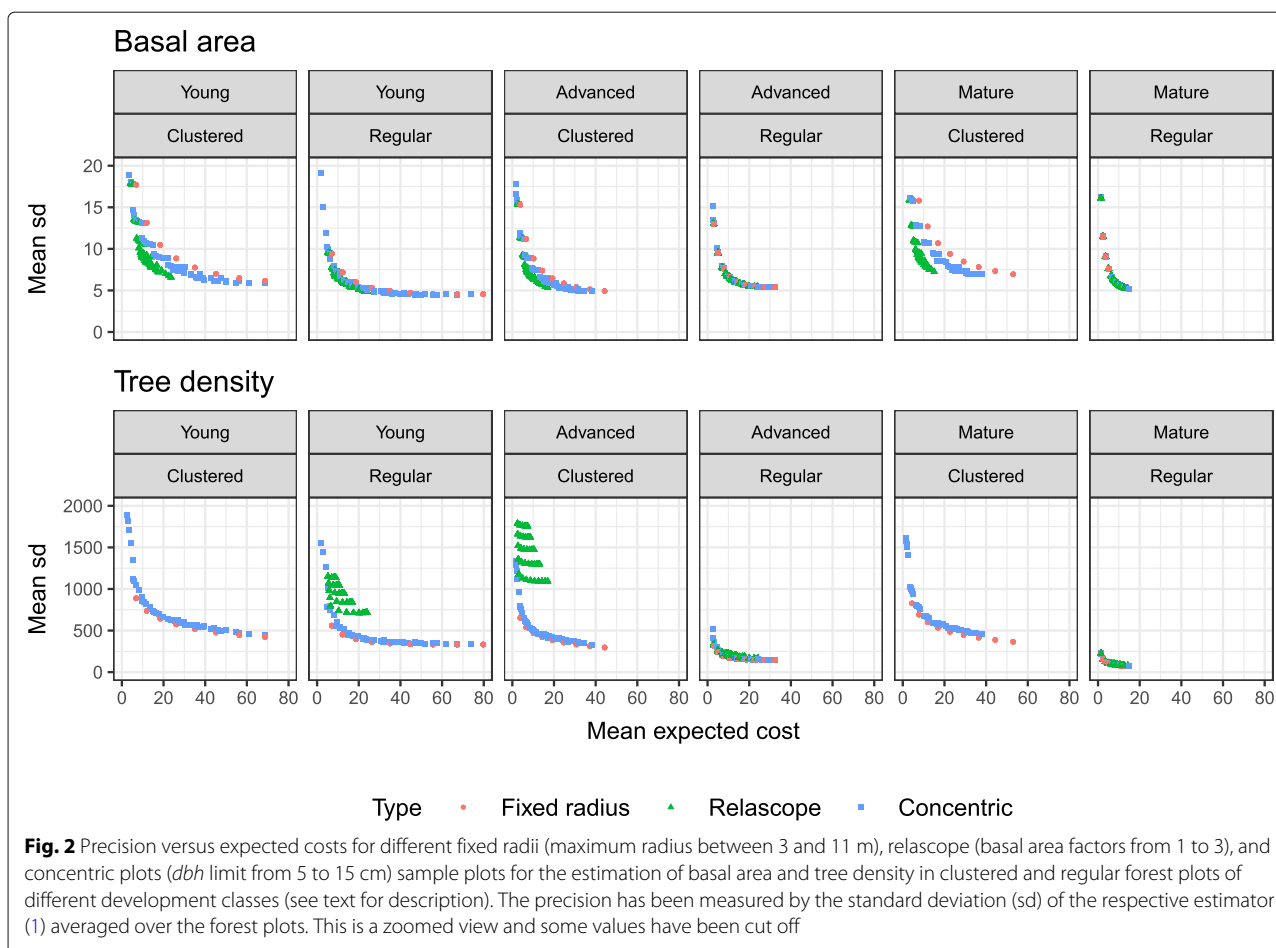
Seedling stands tend to be clustered and forest plots with large basal area regular (Table 2).

The effect of spatial structure on precision and costs

In order to study the effect of clustered and regular patterns on precision and costs of different sample plot types and sizes, we focused on forest plots with medium basal area ([9.8, 32) m^2/ha) not belonging to the seedling class. We selected this group of forest plots as it contains both clustered and regular patterns with more than two observations each. Then, we compared the mean standard deviation of the estimator (1) and mean expected costs (7) per forest structure group and sample plot type and size (Fig. 2). The higher the tree density (younger development classes), the lower the precision of the estimated tree density and the bigger differences between the different sample plot types both for clustered and regular forest plots. The differences between different development classes and different sample plot types in the basal

Table 2 Number of plots per group after classification with respect to development class and basal area (small: $<$ 9.8, medium: [9.8, 32), large: $\geq$ 32 m^2/ha), and spatial structure of forests

Develop. class	Basal area	Structure		
		Clustered	Random	Regular
Seedling	small	2	0	0
	medium	6	1	2
	large	1	0	0
Young	small	0	0	0
	medium	9	0	20
	large	1	0	46
Advanced	small	6	0	1
	medium	9	3	131
	large	0	1	47
Mature	small	4	1	1
	medium	7	5	62
	large	0	1	29
Total		45	12	339

Fig. 2 Precision versus expected costs for different fixed radii (maximum radius between 3 and 11 m), relascope (basal area factors from 1 to 3), and concentric plots (*dbh* limit from 5 to 15 cm) sample plots for the estimation of basal area and tree density in clustered and regular forest plots of different development classes (see text for description). The precision has been measured by the standard deviation (sd) of the respective estimator (1) averaged over the forest plots. This is a zoomed view and some values have been cut off

area estimation were smaller, largest differences between the sample plot types occurring for clustered patterns and in the youngest development class. In any case, both for clustered and regular forest plots, the relascope sample plots appeared to be the most efficient choice to estimate basal area. On the contrary, they were unfeasible for the tree density estimation. In this case, the fixed radius plots seems to be the best design. In general, the precision was lower in clustered forests than in regular forests. Increasing the maximum radius r_{max} (increasing the costs) further led to a more pronounced improvement of precision for clustered than for regular patterns (see also Table 3).

In a more detailed study, we focused on eight different sample plots with two different maximum radii, namely 7 and 9 m. For the relascope plot we considered the basal area factors 1 and 2 for each case. The inner radii of the concentric plots with these maximum radii were 5 and 6 m, respectively, and the *dbh* limit was set to 10 cm. We compared the average expected costs for clustered and regular patterns (Table 3) and found that the fixed radius plots were almost twice as expensive for clustered than for

regular patterns. As expected, the difference in costs was not as prominent for the other designs where fewer small trees are measured. In the case of the relascope plots, it was slightly more expensive to measure the regular patterns. Furthermore, we calculated the cost savings and loss of precision in percentage (determined by the increase in standard deviation of the estimator) in comparison to the values for the concentric plot with a 9 m maximum and a 6 m inner radius, which is the design closest to the one currently used in the Finnish NFI. For clustered patterns, using a fixed radius plot instead of the concentric design would increase the costs by 45% and would achieve a 18% improvement of the tree density estimation, but it would not lead to any improvement of precision for basal area. Again it can be seen that the relascope design has a large loss of precision (over 350%) for the tree density estimation especially for clustered patterns. For the basal area estimation of regular patterns, all designs with a maximum radius of 9 m seem to have about the same precision. When lowering the maximum radius to 7 m, the precision is reduced by 14% on average which comes with a cost saving of between 33 and 52%. The overall best solution for

Table 3 Average expected costs (time in minutes) for measuring eight different sample plots, cost savings (in %), and loss of precision (in %) determined by the increase in standard deviation of the estimator for basal area and tree density

Plot configuration	Exp. cost		Savings %		Precision loss %			
					Basal area		Tree density	
	Clust.	Reg.	Clust.	Reg.	Clustered	Regular	Clustered	Regular
Fixed 7	26	14	10	33	19	12	-5	6
Fixed 9	42	22	-45	-5	0	0	-18	-3
Relas. (1,7)	12	12	59	43	22	13	354	35
Relas. (1,9)	16	18	45	14	5	1	354	29
Relas. (2,7)	9	10	69	52	28	16	523	60
Relas. (2,9)	11	12	62	43	16	8	523	59
Concen. (5,7)	19	13	34	38	20	13	13	10
Concen. (6,9)	29	21	0	0	0	0	0	0

The values are presented separately for clustered and regular patterns. Percentages are calculated in reference to the concentric plot with a 9 m maximum and a 6 m inner radius. For all concentric plots, trees with a *dbh* < 10 cm were only measured in the inner circle

basal area estimation among the evaluated sample plots seems to be the relascope plot with a maximum radius of 9 m and the basal area factor 1, but even the basal area factor 2 gave reasonable precision. For tree density, it appeared worth taking the fixed radius plot with a maximum radius of 7 m into account instead of the concentric plot with a 9 m maximum radius. Not only can costs be saved, but there is even an improvement in precision for clustered patterns.

We also studied the distribution of costs and precision among the clustered and regular forests for the selected eight sample plots with boxplots (Fig. 3). In general, the precision was higher (lower standard deviation of the estimators) for regular patterns than for clustered. Further, the variability was lower for regular than for clustered patterns in all cases, especially for the precision of the relascope estimation of tree density and expected costs of the fixed radius plots. The variability of the basal area estimation appears similar for all sample plots for either clustered or regular forests. Relascope plots had the smallest cost variability, but they showed a large variability in precision for tree density estimation.
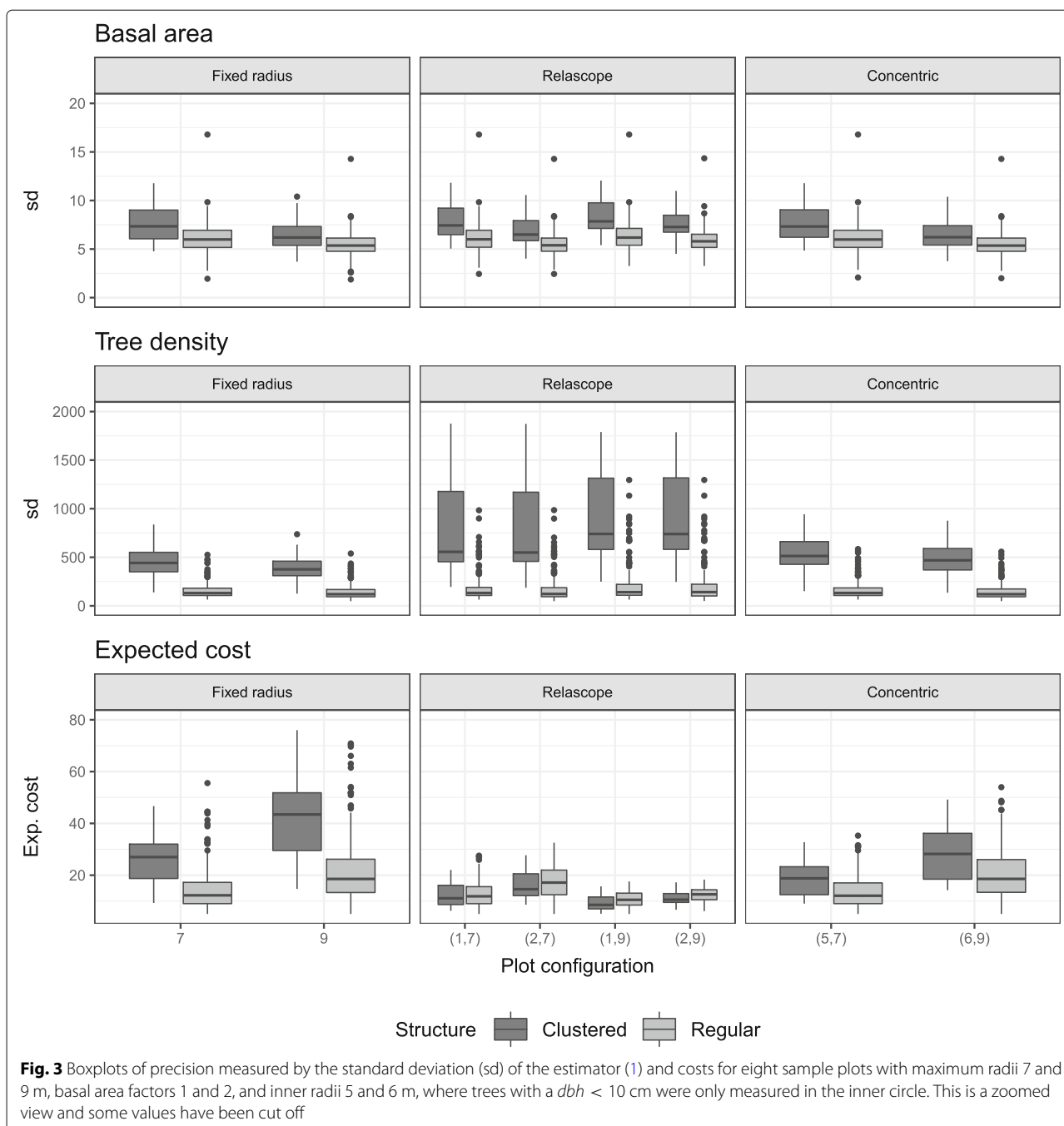
The effect of minimum *dbh* on spatial structure and change estimation

We studied the effect of the minimum *dbh* on the spatial structure classification and the estimation of change in tree density between consecutive time measurements using all 98 measurements from the Erika data source. The spatial structure classification depended on the minimum *dbh* of the trees measured in the forest plots, where with a minimum *dbh* of 0.1 cm, most of the patterns were classified as clustered (Fig. 4). There was a clear shift towards more regular patterns with increasing *dbh* limit.

The minimum *dbh* also had an effect on precision and costs in the change estimation, where with increasing minimum *dbh*, precision increased while cost and cost variability decreased (Fig. 5). Consequently, minimum *dbh*s of 2.5 or 4.5 cm decreased the costs considerably in comparison to the 0.1 cm limit, but the patterns were still clustered (Fig. 4). With a minimum *dbh* of 9.5, the locations of the measured trees mostly formed a regular pattern, where all sample plot types included almost the same trees (Fig. 5). If the main interest is to study the development of young forests, e.g. changes in small trees (with *dbh* ≤ 9.5 cm), the tree density is typically more important than the basal area. In order to minimize the variance (4) of the estimated change such studies should be based on large fixed radius plots (Fig. 5).
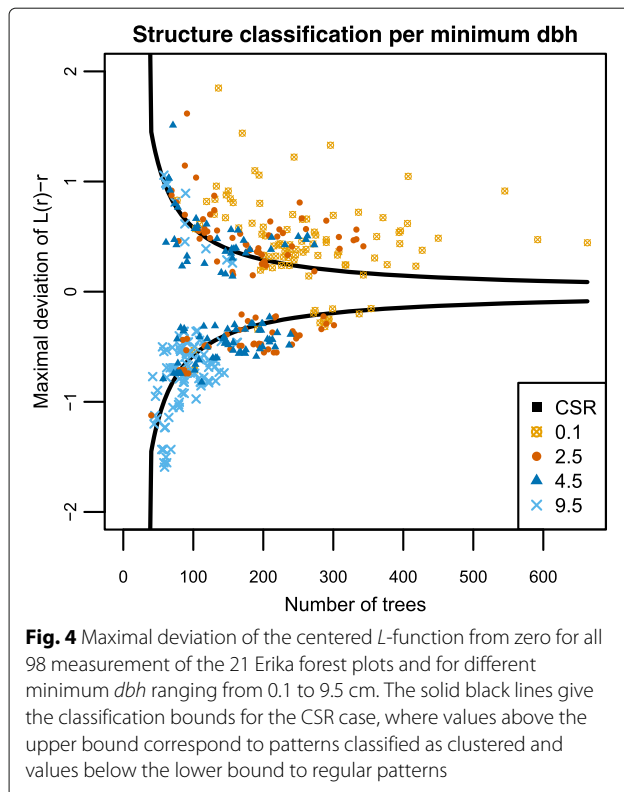
Discussion

It is well known from earlier studies (e.g. Henttonen and Kangas (2015)) that the optimal sample plot type depends on the variable of interest. For volume and basal area, relascope sample plots are both cost efficient and precise, but also other types of sample plots produce precise results. For tree density, fixed area plots are the most precise. This study confirms those results. In cases where the relascope cannot be used reliably, e.g. in very dense forests, additional measurements have to be made to check whether a tree truly belongs to the sample plot. The resulting additional costs have not been included in the cost calculations for this study and could increase the workload variability of relascope plots. We originally considered also more complicated cost models, but finally based the current study on the simple costs (6), because of lack of support for the more complicated cost models. Our simple cost model is based on the assumption that the

Fig. 3 Boxplots of precision measured by the standard deviation (sd) of the estimator (1) and costs for eight sample plots with maximum radii 7 and 9 m, basal area factors 1 and 2, and inner radii 5 and 6 m, where trees with a *dbh* < 10 cm were only measured in the inner circle. This is a zoomed view and some values have been cut off

number of trees and borderline trees are the main factors affecting the costs.

If the spatial structure of forests could be predicted prior to field measurements, for instance from remote sensing data (see e.g. Packalen et al. (2013); Pippuri et al. (2012); Häbel et al. (2018)), it would in principle be possible to let the sample plot size and type vary from a forest to another. However, in the current inventories, the same sample plot is used over the whole inventory region and, thus, a compromise solution is needed. Another aspect is the

operational requirements of NFIs. One solution would be to define a budget for a day's work, and select the sample plot in a cluster by minimizing the RMSEs of the selected forest variables, using the budget as a constraint. However, this approach always leads to measuring as many trees as the budget allows. In addition, it does not take the variability in the workload into account. The plot type with the minimum variability in the costs is the relascope plot and the fixed radius plots have the highest variability due to high variability in the number of measured small trees.

Fig. 4 Maximal deviation of the centered *L*-function from zero for all 98 measurement of the 21 Erika forest plots and for different minimum *dbh* ranging from 0.1 to 9.5 cm. The solid black lines give the classification bounds for the CSR case, where values above the upper bound correspond to patterns classified as clustered and values below the lower bound to regular patterns

For selecting a sample plot for an operational NFI, we need to balance the measurements of small trees so that the RMSE of tree density is acceptable while the plot design is cost-efficient and robust. In order to produce an optimal and operational plot, we could introduce a

probabilistic budget constraint, such that $\mathbb{P}$(time used for measuring a cluster of x plots < 8 h) > p, and minimize the RMSE of tree density subject to that constraint. In this approach, the parameters to be defined are the number of plots within a cluster and the required probability p.

Measurements from forest plots are not only used in inventory calculations, but they are also used as auxiliary variables in model-assisted stand variable estimation. In FMIs, for instance, the collected data are used to build a model to predict the forest variables of interest to all pixels (raster cells) for a map of a certain region of interest. In such a case, the quality of sample plots may be assessed based on the performance of those models. Adnan et al. (2017) studied the effect of sample plot size and stand density on the quality of estimation of tree size heterogeneity from ALS. Tree size heterogeneity can be related to the spatial structure of forests since regular patterns have a tendency to be even-aged with about the same size for all trees and clustered patterns tend to be unevenaged or show at least a larger variation in tree size. Adnan et al. (2017) concluded that for a reliable field-based estimation the smallest sample plot size required is 6 m, but that plot sizes between 9 and 12 m maximize the correlation between field values and ALS metrics. Tomppo et al. (2016) also studied ALS-assisted forest resource estimation, but not for the spatial structure. They suggested that a relascope plot with a basal area factor 1 or a concentric plot with a maximum radius of 9 m and an inner radius of 5.64 m and a lower *dbh* limit of 9.5 cm for the outer circle (used in the Finnish NFI) could be used to

Fig. 5 Precision versus expected costs for different fixed radius (maximum radius between 3 and 11 m), relascope (basal area factors from 1 to 3), and concentric plots (*dbh* limit from 5 to 15 cm) for the estimation of change in tree density (with a certain minimum *dbh* of either 0.1, 2.5, 4.5, or 9.5 cm) between consecutive time measurements for each minimum *dbh*. The precision has been measured by the standard deviation (sd) of the respective estimator (see (4)). The expected cost is the sum of expected costs of the consecutive measurements. Both the mean sd and mean expected cost were obtained by averaging over the forest plots. This is a zoomed view and some values have been cut off

reduce the cost in comparison to a fixed radius plot with a 9 m maximum radius while still being feasible for ALS-assisted inventories. Future studies could assess whether case depended optimal sample plots can be found for remote sensing-assisted inventories.

Conclusions

We classified forests according to spatial structure, tree density and size distribution and found that regular and clustered patterns of trees tend to have different forest characteristics. All in all, it can be concluded that both precision and costs vary depending on the tree density, basal area, diameter distribution, and spatial structure of a forest as well as estimated forest variable. Therefore, there seems to be no obvious overall optimal sample plot.

Today's forests in Finland are most often regular as a result of the silvicultural methods used since the 1950s, which include forest regeneration most often by clear-cut and planting, or due to thinnings from below in younger forests. The amendments to the Forest Act (2013) allow uneven-aged forest management since 2014. This will slowly increase the proportion of clustered forests. This trend makes the monitoring of the development of small trees even more important than before e.g., for the estimation of forest growth and for the prediction of scenarios for the future development of forests. For these purposes, sample plots which give more weight to small trees than relascope plots, where the inclusion probability is proportional to dbh^2, could be considered. For example, this could be achieved by estimating the inclusion probabilities of trees as a function of several tree characteristics (instead of only dbh^2) using data from previous inventories. Also step functions, leading to multiple concentric circles with the radii depending on tree characteristics, could be used in practice.

Author details
[1] Natural Resources Institute Finland (Luke), Latokartanonkaari 9, FI-00790 Helsinki, Finland. [2] Natural Resources Institute Finland (Luke), Yliopistokatu 6, 80100 Joensuu, Finland.

References
Adnan S, Maltamo M, Coomes D, Valbuena R (2017) Effects of plot size, stand density, and scan density on the relationship between airborne laser scanning metrics and the gini coefficient of tree size inequality. Can J Forest Res 47(12):1590–1602

Baddeley A, Rubak E, Turner R (2015) Spatial Point Patterns: Methodology and Applications with R. Chapman and Hall/CRC Press, London

Chiu SN, Stoyan D, Kendall WS, Mecke J (2013) Stochastic Geometry and its Applications. 3rd edn. Wiley, Chichester

Core Team R (2018) R: A Language and Environment for Statistical Computing, Vienna. https://www.R-project.org/

Eerikäinen K, Valkonen S, Saksa T (2014) Ingrowth, survival and height growth of small trees in uneven-aged picea abies stands in southern finland. Forest Ecosystems 1:5. https://doi.org/10.1186/2197-5620-1-5

Forest Act (2013). http://www.finlex.fi/fi/laki/kaannokset/1996/en19961093.pdf

Häbel H, Balázs A, Myllymäki M (2018) Spatial analysis of airborne laser scanning point clouds for predicting forest variables. arXiv:1805.08907 [stat.AP]., https://arxiv.org/abs/1805.08907

Henttonen HM, Kangas A (2015) Optimal plot design in a multipurpose forest inventory. Forest Ecosystems 2(1):1–14. https://doi.org/10.1186/s40663-015-0055-2

Hökkä H, Koivusalo H, Ahti E, Nieminen M, Laine J, Saarinen M, Laurén A, Alm J, Nikinmaa E, Klöve B, Marttila H (2008) Effects of tree stand transpiration and interception on site water balance in drained peatlands: experimental design and measurements. In: Farrell C, Feehan J (eds). After Wise Use - The Future of Peatlands, Proceedings of the 13th International Peat Congress, Tullamore, vol. 2. pp 169–171

Horvitz DG, Thompson DJ (1952) A generalization of sampling without replacement from a finite universe. J Am Stat Assoc 47(260):663–685. https://doi.org/10.1080/01621459.1952.10483446

Kangas A, Astrup R, Breidenbach J, Fridman J, Gobakken T, Korhonen KT, Maltamo M, Nilsson M, Nord-Larsen T, Næsset E, Olsson H (2018) Remote sensing and forest inventories in nordic countries - roadmap for the future. Scand J Forest Res 33(4). https://doi.org/10.1080/02827581.2017.1416666

Kreutz A, Aakala T, Grenfell R, Kuuluvainen T (2015) Spatial tree community structure in three stands across a forest succession gradient in northern boreal fennoscandia. Silva Fenn 49(2):397–412. https://doi.org/10.14214/sf.1279

Mäkinen H, Isomäki A (2004) Thinning intensity and growth of scots pine stands in finland. Forest Ecol Manag 201(2–3):311–325. http://dx.doi.org/10.1016/j.foreco.2004.07.016

Mandallaz D (2007) Sampling Techniques for Forest Inventories. CRC Press, Boca Raton

Mandallaz D, Ye T (1999) Forest inventory with optimal two-phase, two-stage sampling schemes based on the anticipated variance. Scand J Forest Res 29(11):1691–1708

Næsset E (2004) Practical large-scale forest stand inventory using a small-footprint airborne scanning laser. Scand J Forest Res 19(2):164–179. https://doi.org/10.1080/02827580310019257

Packalen P, Vauhkonen J, Kallio E, Peuhkurinen J, Pitkänen J, Pippuri I, Strunk J, Maltamo M (2013) Predicting the spatial pattern of trees by airborne laser scanning. Int J Remote Sens 34(14):5154–5165. https://doi.org/10.1080/01431161.2013.787501

Päivinen R (1987) Metsän inventoinnin suunnittelumalli. [A planning model for forest inventory, In Finnish]. 11th edn. University of Joensuu publications in Sciences, University of Joensuu, Joensuu

Pippuri I, Kallio E, Maltamo M, Peltola H, Packalén P (2012) Exploring horizontal area-based metrics to discriminate the spatial pattern of trees and need for first thinning using airborne laser scanning. https://doi.org/10.1093/forestry/cps005

Therneau T, Atkinson B, Ripley B (2017) rpart: Recursive Partitioning and Regression Trees., https://CRAN.R-project.org/package=rpart, r package version 4.1-11

Tomppo E (1986) Models and methods for analysing spatial patterns of trees. Communicationes Instituti Forestalis Fenniae 138

Tomppo E, Gschwantner T, Lawrence M, McRoberts RE (eds) (2010) National Forest Inventories. Pathways for Common Reporting. Springer, Heidelberg

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, Dordrecht

Tomppo E, Kuusinen N, Mäkisara K, Katila M, McRoberts RE (2016) Effects of field plot configurations on the uncertainties of ALS-assisted forest resource estimates. Scand J Forest Res 32(6):488–500. https://doi.org/10.1080/02827581.2016.1259425

Zeide B (1980) Plot size optimization. Forest Sci 26(2):251–257