SIMULATING SAMPLING EFFICIENCY IN AIRBORNE LASER BASED FOREST INVENTORY

SIMULERING AV SAMPLING EFFEKTIVITET I SKOGKARTLEGGING VED FLYBÅREN LASERSKANNING

LIVIU THEODOR ENE

NORWEGIAN UNIVERSITY OF LIFE SCIENCES
DEPARTMENT OF ECOLOGY AND NATURAL RESOURCES MANAGEMENT
MASTER THESIS 30 CREDITS 2007

Abstract

A simple simulator was developed to test whether airborne laser scanning can be used in forest inventories as a strip sampling tool for forest inventories purposes. The simulator was based on the existing two-stage, grid-based sampling procedure. A forest stand simulator and laser-derived single tree models were used to create a virtual forest area. Sampling simulations using different designs were run, and then laser scanning-based mean volumes and ground-plots based estimates were integrated using Monte Carlo technique. The simulation results were then assessed against the truth population value. Multiplicative regression models and ground-plot based inventory gave unbiased mean volume estimates. The lowest RMSE for regression-based Monte Carlo estimates was 5.1 m^3/ha (2.0%) and the highest was 8.4 m^3/ha (3.3%). The RMSE for the ground-plot Monte Carlo estimates varied between 13.7 m^3/ha (5.4%) and 18.4 m^3/ha (7.2%). Relative efficiency of laser-based estimates was from 1.8 and up to 3.0 times higher compared to estimates based on ground-plot inventory. The results indicated that forest surveys over large areas carried out using airborne laser scanning as a strip sampling tool can provide accurate estimates, and can be more effective than traditional systematic ground-plot based inventories.

Keywords: forest inventory, scanning LIDAR, simulator

Sammendrag

En enkel simulator ble utviklet for å teste om flybåren laserskanning kan tas i bruk som samplingsverktøy for skogtaksering. Simulatoren er basert på den eksisterende, tostegs grid-baserte samplingsmetoden. En skogbestandssimulator og laserdata fra faktiske skogbestand ble brukt til å skape et virtuelt skogområde. Samplingssimuleringer med forskjellige design ble utført ved hjelp av Monte Carlo teknikk og de gjennomsnittlige volumestimatene fra laserskanningsmetoden og fra tradisjonell prøveflatetakst ble beregnet. Simuleringsresultatene ble sammenlignet med den sanne populasjonsverdien. Multiplikative regresjonsmodeller og prøveflatetakst gav forventningsrette volumestimater. Den laveste RMSE ved bruk av laserskanningsmetoden var på 5.1 m^3/ha (2.0%), og den høyeste verdien var på 8.4 m^3/ha (3.3%). RMSE for prøveflatetakst varierte fra 13.7 m^3/ha (5.4%) till 18.4 m^3/ha (7.2%). Den relative effektiviteten til de laserbaserte estimat var fra 1.8 til 3.0 ganger større enn estimat basert på prøveflatetakst. Resultatene indikerte at samplingsbasert laserskanning over store skogområder kan gi nøyaktige estimat, og at laserskanning kan være mer effektiv enn den tradisjonelle systematiske prøveflatetakst.

Nøkkelord: skogtakst, flybåren laserskanning, simulator

Content

1. Introduction

Forest inventories have been dynamically changing in scope and objectives. The change has been driven by various interpretations of the concept of sustainability. Assessment of additional environmental characteristics has gradually been included as a part of forest inventories over time, and nowadays national forest inventories should be able to provide also reliable biomass/carbon estimates to meet the requirements of the United Nations framework convention for climate change.

During the past two decades, remote sensing techniques have proven to be able to meet some of the demand for environmental related data at fairly low cost. Among these techniques, small footprint LIDAR data (Light Detection and Ranging) has become one of the most common remotely sensed data sources for analysing the canopy structure at the scale of operational forest management (Wynne 2006).

It has been shown that profiling LIDAR can provide reliable biomass sampling-based estimates at low costs (Nelson et al., 2006). The procedure consists of a two-stage sampling scheme. First, profiling transects are taken by flying parallel fight lines separated by many kilometres over the area in question. Systematically distributed ground plots or ground transects are measured along the LIDAR transect. Ground-based estimates are regressed against LIDAR measurements, and the resulting regression equation are used for prediction along the LIDAR transects across the entire sampled area. Up to now, it has been difficult to accurately co-register the LIDAR data and the ground observations because precise geolocation of the LIDAR data has been lacking. For the ground observations, only simulated LIDAR data has been used in practice. This may induce biased LIDAR-based estimates (Nelson et al., 2004).

Commercial airborne laser scanners avoid this problem, since flight lines corresponding to laser scanning strips provide an accurately located cloud of 3-dimensional (3D) observations, which easily can be related to ground measurements such as plots of various shapes and sizes.

Scanning LIDAR is today used operationally for stand-based “wall to wall” inventories of forest stands in Norway (Næsset 2004). However, for larger areas such as counties or country regions, “wall to wall” inventories are not feasible and so far only profiling lasers have been used for large area biomass surveys (Nelson et al., 2004). The alternative to reduce the costs is to use laser scanning in a sampling mode, by flying parallel, equally spaced strips over the study area and collecting the ground-truth references only within the sampled strips, using systematic sampling schemes.

When sampling-based forest inventory systems are designed, it is important to find an optimal balance between ground sampling efforts and amount and density of LIDAR data acquired over the area in question. Because sampling applications often are relevant in areas with a size where it is not feasible to establish a ground-truth reference value, designing optimal inventory systems has to rely on some kind of simulation where different combinations of field and airborne data collection can be explored.

The aim of this master project was to develop a simple small-scale simulator for laser scanning-based strip sampling design for timber volume assessment, based on the two-stage, grid-based sampling procedure developed and tested by Næsset & Bjerknes (2001) and Næsset (2002, 2004), and to assess the relative efficiency of regression-based laser scanning estimates relative to the correspondent ground-plot based estimates, under various sampling designs.

2. Materials

2.1 Stand data

Two datasets containing forest stand parameters derived from ground measurements were used as empirical input data for simulations. Further details presented for these datasets can be found in Næsset (2004) and Bollandsås & Næsset (2007).

The first dataset contains ground data collected in summer 2003. These measurements were used as model calibration data in two previous studies (Bollandsås & Næsset 2007, Solberg et al., 2006). Twenty circular plots of 0.1 ha were collected from a boreal nature reserve located in south-eastern Norway. The forest was multilayered, with a broad range of tree sizes and stand ages, and dominated by Norway spruce [*Picea abies* (L.) Karst.] and Scots pine (*Pinus silvestris* L.). The plots were establish in subjectively selected spruce dominated sites.

On each plot, all trees with diameter at breast height ($d_{bh)}$ $\geq$ 3cm were callipered and tree heights were measured on trees selected with probability proportional to stem basal area. Mean diameter was defined as diameter corresponding to mean stem basal area (d_{BA}), and means height was defined as the average basal area weighted (Lorey's) height (h_L), see Table 1.

For all trees with d_{bh} > 3 cm, the polar coordinates from the plot centre were registered. Height, height of crown base, crown radius in four cardinal directions and average crown diameter were measured on trees selected from each plot.

Both Global Positioning System (GPS) and Global navigation Satellite System (GLONASS) were used to determine the planimetric plot coordinates (Euref89). A Topcon Legacy GPS + GLONASS receiver, observing both the pseudo range and carrier phase, were used as rover receiver with a setup of 2 seconds logging-rate and 15˚ cut-off angle. On each plot, logging period ranged between 0.5 to 1.5 hours, with an average antenna height of 4 m. The base station, a similar Topcon Legacy GPS + GLONASS receiver, was established within a distance < 2.5 km from the sample plots, with an accuracy of 0.4 cm of planimetric coordinates. During post-processing, the records from the base station were used as reference for the rover coordinates. The cut-off angle for base station was set to 12˚, in order to ensure that the rover and the base station receive signals from the same satellites. The average estimated accuracy of the plot coordinates was 10 cm.

Table 1: Summary of selected 0.1ha plot-derived stand parameters for Norway spruce from the first dataset

	d_{BA}	BA	N	$d$max	h_L
max	30,1	46,4	1040	60,6	28,9
min	19,9	28	650	39,6	17,7
mean	17,19	27,42	670,6	37,3	16,99

Note: d_{BA} = basal area mean diameter (cm); h_L = basal area weighted mean height (m); BA = basal area (m^2/ha); N = stem number; d_{max} = maximum diameter (cm).

The second dataset comprised 60 large plots (see Næsset 2004). Data were collected in summer 2001 from a productive forest area of approximately 5000 ha located in the municipality of Krødsherad, south-east Norway. The forest composition was dominated by Norway spruce and Scots pine, while younger stands were dominated by deciduous species, mainly birch (*Betula pubescens* Ehrh.). Stratification in three predefined strata (young forest, mature forest on poor sites, and mature forest on good sites) according to site quality was carried out over all plots (Table 2).

The plots were supposed to be quadratic (61x61 m), and differential GPS + GLONASS were used to determine the position of each corner for all plots. Dual frequency Javad Legacy GPS + GLONASS receiver observing pseudo range and carrier phase was used as rover, with a logging rate of 2 sec and antenna height ranging between 1.8 to 3.6 m. Logging periods ranged between 0.2 to 1.2 hours. An identical receiver was established as base station, at a distance <10 km from the plots. The average accuracy of the planimetric plot corner coordinates was reported to range from <0.1 to 1.4 m, with an average of approximately 0.1 m. Due to practical impediments, the plot form was deviating from the quadratic, which gave a horizontal plot area from 3121 to 4219 m^2, with an average of 3736 m^2.

Within each plot, all trees with diameter at breast height $d_{bh} \geq$ 4cm and $d_{bh} \geq$ 10cm were callipered in young and mature stands, respectively, using 2 cm diameter classes. Heights measurements were taken from trees selected with probability proportional to stem basal area at breast height. For each plot, the mean height corresponding to Lorey's height was computed form the mean height of the individual diameter classes, weighted by total plot basal area for each diameter class.

Table 2: Summary of selected large plots derived stand parameter from the second dataset:

	d_{BA} *g*	d_{max} *g*	h_L *g*	*Ng*	d_{BA} *f*	d_{max} *f*	h_L *f*	*Nf*	d_{BA} *l*	d_{max} *l*	h_L *l*	*Nl*
max	27,6	51	22,2	904	36,3	49	23,6	622	23,9	47	20	286
min	12	17	8,4	13	18,7	35	12,7	10	11,5	13	10,8	5
mean	17,4	32,1	14,9	359,5	21,6	35,5	14,4	228,5	12,6	21,7	10,6	81,5

Note: d_{BA} = basal area mean diameter (m^2/ha); d_{max} = maximum diameter (cm); h_L = basal area weighted mean height (m); *N* = stem number; *g* = Norway spruce; *f* = Scots pine; *l* = deciduous trees (assimilated with birch)

The program package SILVA2.2 (Pretzsch et al., 2002) was employed to create the virtual forest. The result of simulation with SILVA2.2 stand generator is a single tree description concerning tree species, diameter at breast height (d_{bh}), total height (h), and height of crown base(h_{cb}), crown diameter (c_{d}), and tree coordinates (x, y) for each forest stand. The tree key variables for initialization of a stand simulation are tree species, d_{bh} (cm), d_{max} (cm), h_{L} (m), and N/ha, while age and basal area (G, ha/m^2) are not mandatory.

Not all the plot-derived stand parameters from the field-data could be used as input parameters for SILVA 2.2 forest stand generator, probably due to some inadvertencies between test plot reference data and SILVA2.2 model calibration. From a total of 80 field plots, 13 plots from the first dataset and 25 from the second one gave satisfactory results when they were used to generate stands with SILVA 2.2, and only these were considered further.

2.2 Laser data

Laser scanner data from the same area as the first dataset presented above were acquired during June 2005 (leaf-on canopy condition) with an Optech ALTM 3100 sensor operating at 100 kHz laser pulse repetition rate and 70 Hz scanning frequency. The plane was flown approximately 750 m above ground with an average speed of 75 m/s. The maximum scan angle was 20°, and the corresponding swath width was about 264 m. Pulses transmitted at scan angles that exceeded 8° were excluded from the final dataset. The average footprint size was about of 21 cm, with an average density of 5.09 points/m^2. First and last echo were recorded.

2.3 Laser-derived single tree models

Laser pulses and ground measurements collected in summer 2003 from 0.1 ha stand plots, as described above (Table 1), were used to derive individual-tree models. Each laser pulse was related to a tree crown projection by the mean of planimetric coordinates, and then each of the resulted laser point clouds was considered to represent a spatial crown model for Norway spruce.

Single-tree models were build using field and laser data. A tree-model was represented as a unique combination of diameter, height, crown height, crown projection radius, laser pulse heights and volume values (Table 3). In this way, a number of 435 models of Norway spruce trees were created and used to populate the virtual study area.

For each tree-model, the volume was calculated by the means of functions for Norway spruce with bark (Vestjordet 1967). These functions are currently used by the Norwegian national forest inventory and are supposed to give a standard deviation of 8 to 10% of the volume.

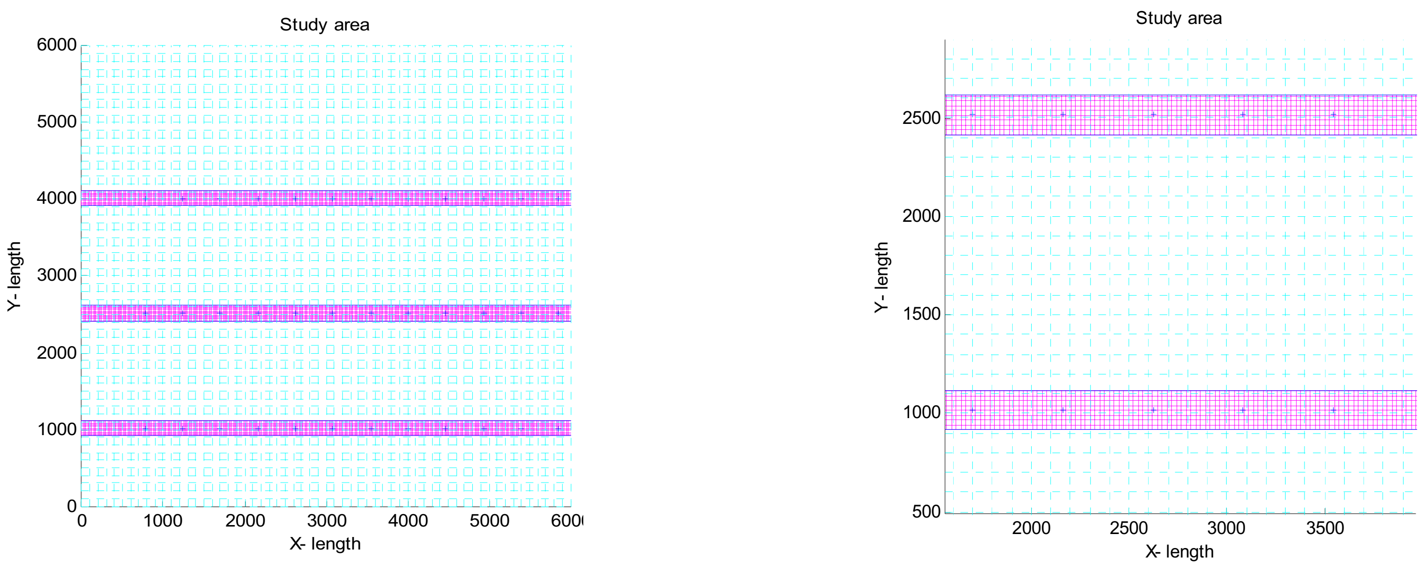
Table 3: Descriptive statistics for individual tree-model parameters:

Variable	mean	st.dev	min	median	max	IQ
d_{BA} (cm)	19,8	10,42	3,2	18,8	51,0	16,8
h_L (m)	15,8	6,1	3,6	16,0	29,5	9,9
crown height(m)	3,4	2,35	0,2	3,1	13,5	3,6
crown radius(m)	1,3	0,4 2	0,6	1,3	2,9	0,63
volume (m^3)	0,38	0,41	0,003	0,22	2,46	0,54

3. Methods

The strip sampling simulation method was based on the two-sampling procedure described by Næsset & Bjerknes (2001) and Næsset (2002, 2004). In the first stage, georeferenced sample plots centred along the strips in a systematic design were used to regress ground-based estimates against laser-derived metrics. In the second stage, a regular quadratic grid was used to divide each strip into cells and laser-derived metrics were derived from each grid cell (Figure 1). Regression equations obtained in the first stage were then used for prediction at cell level. The sampling unit is the strip, but because all strips had equal dimensions and contained the same number of grid cells, the total volume was estimated as the sum of predicted volume for all grid cells over all strips.

In parallel, estimation of mean volume by ground-plots systematic sampling was also done, as a kind of conventional sampling-based inventory.



a. three equally spaced strips of laser data divided in grid cells;

b. details: the dark dots represent the centres of systematically distributed ground – plots;

Fig. 1: Laser scanning- based strip sampling design for a non- stratified population

Estimates of population mean volume and its sampling error were derived by Monte Carlo (MC) integration over 50 simulation results, and then bias, standard deviation and RMSE for estimated mean values were used for assessment of sampling designs and methods against the reference volume of a predefined population model. Relative efficiency of regression-based estimates obtained from laser scanning strip sampling and ground-based systematic plot sampling estimates was assessed for each sampling schemes.

3.1 Population model

For simplicity, the study area was considered flat and the population was considered free of any trends. Furthermore, neighbourhood effects were ignored as well. In this way, the spatial structure in each cell supposed to be independent of the position in the array.

The 2D-study area was then defined in a local coordinate system by choosing the lengths on two Cartesian axes, in 100m units. The result was a rectangular polygon, with (x, y) lengths multiple of 100m, because this length unit was considered to be appropriate for the scale of the present project.

The frame of study area was considered as an array. To create the population, at each (i, j) array position was randomly allocated one of the 38 forest stand models of 1.0 ha generated by means of SILVA 2.2. In each cell, each tree got a pair of (x, y) local coordinates.

In the next step, the (x, y)-locations of each tree in a cell were related to one of the single-tree models derived from laser scanning data. In this way, the program substituted the virtual trees created by the stand generator with empirical tree models based on laser scanning measurements, and the resulted spatial structure could also reflect the 3D-competition relationships within each virtual forest stand.

For other species than Norway spruce present in the second dataset (Table 2), there were no available data to build individual laser-based tree models. For diameter matching and volume calculations, these trees were assimilated to Norway spruce, but their (x, y)-location was kept unchanged.

Because only 435 single-tree models could be derived, tree breast height diameter was used as a key to join laser-derived tree models to trees generated by SILVA 2.2. An algorithm matched all trees in the study area with diameters among the laser-based single-tree models. Then, the single-tree parameters derived from laser scanning measurements were transferred to the corresponding diameter-equivalent trees positioned at (x, y)-coordinates in the study area.

The matching results often consisted of several trees with equal diameters. Since a further search based for instance on height and/or crown height was not performed from the reason mentioned above, the algorithm was designed to select randomly only one tree-model among all trees models with the same diameter, and to replace the tree at the position (x_i, y_i) from the generated forest stand with the laser- derived tree model. The choice of using randomly selected trees and not for instance the first found one may add some variation within each stand and consequently across the study area. If diameter matching didn't occur, the tree with diameter closest to the desired value considering either larger or smaller diameters, was selected instead.

When single tree models are located by diameter matching to the corresponding (x_i, y_i) position in the study area, the rest of the tree parameters (height, crown height, crown radius, laser pulse heights, and single-tree volumes) are also related to the same (x_i, y_i) locations. In this way, study area was populated with laser derived tree models, and the volume of the entire population could be calculated as the sum of individual trees.

The laser scanning data consists of clouds of laser pulses related to tree crowns. Each laser pulse, representing the first echo, has known (x, y and z) coordinates. In this analysis, the (x, y) coordinates of each laser point were discarded. It was assumed that all laser points related to one tree inside a cell fall inside the same cell and that all pulses inside a tree crown projection belong only to that tree.

The dataset used in this project did not include ground echoes. To calculate laser-derived metrics as canopy density, it was considered that each cell has got a uniform coverage of laser points, so that the total number of pulses within a grid cell could be linearly extrapolated from the number of pulses that fall inside crown polygons. Pulses with heights below 2 m were also considered as ground points.

3.2 Simulation process

The simulation program was written using Matlab 'The Mathworks, Inc.', an interpretative user-friendly language which can perform computationally intensive tasks faster than many other traditional programming languages. The program has a top-down, structured design and it contains three main modules, as shown in Appendix A. At each simulation start, the internal state of the chosen Matlab random number generator is set to a fixed state so that the simulations results can be repeated.

First input parameters (X, Y-lengths) define the size of study area. Next inputs are laser sampling design parameters: no. of laser flight strips, strip width, distance among strips, and parameters for field sampling design: number of plots, plot area and distance among plots (Table 4). Finally, the number of iterations for model selection and sampling are requested. Basic error checking procedures ensure that the choices are compatible with the study area and with each other.

Based on previous findings (e.g. Magnussen et al., 1998; Næsset 1997, 2002, 2004), two independent variables derived form first laser pulse returns were considered for volume prediction for each grid cell: canopy density, defined as proportion of first pulse laser hits to total number of laser hits in each grid cell, and a quantile of the laser height distribution within the same grid cell. The relationship between sample quantiles of laser data and the canopy attributes has been described in previous studies (e.g. Magnussen and Boudewyn, 1998). The percentile corresponding to the 9^{th} quantile of laser canopy height (h_{90}), together with the canopy density corresponding to the proportion of the first pulse laser hits (d_0) were then considered as candidate regressors.

Trials of exploratory regression analysis were carried out before simulation to detect possible deviations from model assumptions and to test procedures for coping with them. During simulations, sampling outcomes may require different variance stabilizing transformations. Anyway, when the aim is to assess the performance of a regression model, only one transformation should be applied to data during all simulations. Different outcomes of the dependent variable (sample plot volume) were analytically assessed by the means of Box-Cox transformation method.

Five regression models were then proposed: a multiplicative model, a linear model without transformations, and three different models with transformed response variable: log(y), sqrt(y) and asin(sqrt(y)). For the multiplicative model, only two independent variables (h_{90} and d_0) were used, and consequently this model was not subject to stepwise selection.

For all the other models, an empirical approach was used to find a "most suitable" regression model form for a given sequence of samples. First, a random stripe sampling scheme was generated. The location of each stripe and its correspondent sample plots were hold fixed, and several outcomes of possible study area were generated and sampled. During iterations, stepwise regression was used for model selection, and each subset model was registered. After running all iterations, the most frequently used model form for each regression model was selected as a final model to be used for sampling simulations.

The subset models resulted from stepwise regression were compared with alternative models derived with best subsets selection procedure, using adjusted-R^2 and C_p-statistic criteria focused on selection of unbiased regression models. It is desired that Cp is small and close to p, where p is the number of coefficients (i.e., number of variables + 1 for models including an intercept). A value of Cp equal to p + 1 suggests that the model contains no estimated bias (Mendenhall & Sincich, 2003).

It was considered necessary to assess stepwise regression against best subsets selection method, since large VIF's were expected to occur due to inclusion of second-order terms as independent variables. VIF's larger than 10 imply serious problems with multicollinearity and indicates that the associated regression coefficients are poorly estimated and theirs values are very sensitive to data form the particular sample used for estimation, and therefore very unstable to relatively small changes in the data points. Even though variable selection method such as stepwise selection can be applied, multicollinearity can seriously affect the performance of the method. Because the purpose of regression models was supposed to be estimation and prediction, and not to establish the cause and effect relationships, all independent variables were kept in the model. However, stepwise regression was an alternative to be investigated especially because the algorithm can easily be implemented into simulation software.

The number of iterations must be known before to run the simulation. Initial tests has showed that after a number of ca 50 iterations, both regression and sample plot based MC estimates of mean timber volume seem to converge to stabile values. However, the number of iterations should vary with the study area, sample design and population variability.

The simulations were run for the sampling schemes presented in Table 4. The study area was defined as a quadrate with an area of 3600 km^2. Sample plots of 200, 400 and 600 m^2 were used to provide ground estimates. Using quadratic plots avoids distance calculations for allocating trees into each plot, and significantly improves the computational performance during simulation. Parallel laser strips with widths of 160, 180 and 200m and spacing of 1500 m were generated. This is in accordance with previous studies related to profiling laser sampling (Nelson et al 2006), which assert that the distance between sampling strips systematically allocated should be ≤ 4 km for sampled areas between 1000 to 5000 km^2. The sampling intensity for different plot sizes was hold almost constant, but sample size varied with plots sizes (Table 4). Compared to sampling intensities in ongoing research studies, which typically are less than 0.003% (Gobakken et al., 2006), the sampling intensity at stand plots level is obviously much higher, but necessary to reach ground samples large enough to obtain reliable regression estimates.

Table 4: Sampling schemes

Sampling intensity (%)		Plot area	Number of plots		Plot spacing	Strip spacing	Strip width	Strip length	Strip area (ha)	
strip	plots	(m^2)	per strip	total	(m)	(m)	(m)	(m)	unit	total
	0,57	200	34	102	172					
8	0,57	400	17	51	334	1500	160	6000	96	288
	0,60	600	12	36	462					
	0,57	200	34	102	172					
9	0,57	400	17	51	334	1500	180	6000	108	324
	0,60	600	12	36	462					
	0,57	200	34	102	172					
10	0,57	400	17	51	334	1500	200	6000	120	360
	0,60	600	12	36	462					

Finally, the MC estimates for both laser stripe and ground-based systematic sampling were assessed by the means of a two tailed t-test against the population value. Bias, standard deviation and RMSE for MC estimates of mean volume were then used to assess the sampling designs and regression models. Relative efficiency of regression-based laser scanning estimates relative to correspondent ground-based estimates was calculated as the ratio of theirs RMSE.

4. Results

4.1 Exploratory analysis

Several samples were generated for all sampling schemes in order to be analysed before running the simulations. For instance, for sampling schemes using plots of 200 m^2 and strips spaced at 160 m, the residual plots for the main effect model presented a curvature pattern (Appendix B-Figure 4.a). Assessment of the full quadratic model (Appendix B-Figure 4.c) indicated that variance stabilizing transformations might be helpful to correct model inadequacy. The results of Box-Cox transformation (Appendix B-Figure 4.b) varied, indicating either square root or log-transformation of the dependent variable, but, at least for the samples which were analyzed, the square root transformation occurred more frequently($\lambda = 0.5$). This result was somehow unexpected, since multiplicative regression models in logarithmic variables are traditionally used to predict laser-scanning derived parameters of interest(e.g. mean height, dominant height, mean diameter, stem number, basal area and volume). Residuals for stepwise subsets (Appendix B-Figure 4.c-g) have shown that subset models seemed to follow the model assumptions.

For plots of 200 and 400 m^2, stepwise procedure and best subsets models yield nearly similar values for adjusted-R^2, prediction-R^2 and PRESS statistic, while for plots of 600 m^2 the best subsets selection gave models with slightly higher values of adjusted-R2, prediction-R2 and lower values for PRESS statistic. In comparison to stepwise regression, best subsets method produced models with slightly higher values of adjusted-R^2, prediction-R^2 and lower PRESS statistic (Table 5). Mallow's Cp values for subsets selected by stepwise regression were usually deviating from the value p = number of variables + 1, more than the values of Cp-criterion for best subsets method. For plots of 600 m^2, the difference between Cp-values and p+1 were the highest, for all regression models. Furthermore, it was noticed that stepwise regression may produce subsets that are not among selected models by the best subsets procedure, because of a very large value of Cp-criterion.

In the case of multiplicative model, for the same samples, adjusted-R^2 and prediction-R^2 yield values of 91.0 and 90.58% for plots of 200 m^2, and values of 90.8 and 89.72% for plots of 400 m^2. The performance of these criteria decreased to 86.8 and 86.16% respectively, for plots of 600 m^2.

Tab. 5: Comparison between stepwise regression and best subsets derived models

Model	Stepwise regression					Best subsets				
	Regr.coeff	R^2_{adj}	C_p	R^2_{pred}	PRESS	Regr.coeff	R^2_{adj}	C_p	R^2_{pred}	PRESS
					plot area 200m2, strip width 160m, λ = 0.5 iteration 1					
2	b1 b2 b3	90,2	2,5	89,41	2,33	b1 b2 b4	90,1	3,3	89,36	2,34
3[a]	b2 b5	90,8	3,9	90,48	2,37	b2 b5	90,8	3,9	90,48	2,37
4[b]	b3 b5	89	11,3	88,13	2,06	b1 b4	89	3,1	88,17	2,06
5	b1 b5	91,5	1,1	91,09	47,28	b4 b5	91,3	3,4	90,89	48,37
					plot area 400m2, strip width 160m, λ = 0.5, iteration 1					
2	b5	91,3	3,9	90,7	0,68	b1 b3	91,6	3	90,82	0,69
3[a]	b5	92,7	0,1	92,31	1,44	b5	92,7	0,1	92,31	1,44
4	b5	89,5	3,7	88,61	2,23	b1 b5	89,9	2,7	88,67	2,22
5	b5	92,5	4,5	92,05	64,32	b1 b2 b5	93,1	2,3	92,6	59,91
					plot area 600m2, strip width 160m, λ= 0.5, iteration 1					
2	b5	85,8	10,1	84,4	0,39	b1 b3 b4	88,1	5,1	86,53	0,34
3	b5	86,8	5,8	85,4	1,37	b3 b4 b5	88,3	3,7	86,31	1,28
4	b5	82,3	6,7	79,13	2,42	b1 b3 b4	84,2	4,4	81,09	2,19
5	b5	86,2	5,9	84,24	93,35	b1 b3 b4	87,6	3,8	85,57	85,51

Regression coefficients: b1 - canopy density; b2 - (canopy density) ^2; b3 - height percentile; b4 - (height percentile) ^2; b5 – interaction term (canopy density x height percentile);
Models: 2 – log(y) ; 3 – sqrt(y); 4 – asin(sqrt(y)); 5 – linear;
Note: [a] final subset model for both procedures; [b] stepwise regression model was not included as outcome for best subsets

4.2 Simulation results

With the exception of multiplicative model, the final regression equations were build using stepwise regression ($p_{in} = 0.05$, $p_{out} = 0.10$). Before each simulation, a number of 20 iterations were used to select final regression models. Each simulation included 50 sampling-and-prediction iterations.

A number of 45 mean volume estimates and theirs RMSE were derived using five regression models (Table 6). In addition, for each sampling scheme, an estimate of mean volume and the correspondent RMSE were derived by ground-based systematic plot sampling (Table 7). The reference value of mean volume per ha was of 254 m^3 i.e. total population volume of 914,400m^3 divided by the size of the study area of 3600 ha. The simulated study area included over 2.7 million trees. The number of iterations used for each simulation ensured convergence for both regression and ground-plot based estimates (Appendix C). For mean timber volume estimates, the convergence occurred after ca 40-45 iteration for sampling schemes using ground plots of 200 m^2, ca 20-30 iterations for plots of 400 m^2 and after ca 15-25 iterations for plots of 600 m^2. Sampling error approached asymptotically the standard deviation of mean volume estimates resulted from simulations (Appendix C).

Tab.6: Bias, standard error and RMSE for regression estimates of mean volume (m³/ha)

Strip width (m)	Model	Plot area 200 m² bias	std.error	RMSE	400 m² bias	std.error	RMSE	600 m² bias	std.error	RMSE
160	1	-0,8 ns	5,5	5,6	-0,8 ns	5,7	5,8	-0,2 ns	6,4	6,4
	2	-2,2*	6,2	6,6	-1,3 ns	6,4	6,5	-0,8 ns	7,4	7,4
	3	-1,5*	5,2	5,4	-1,6 ns	5,7	5,9	-0,9 ns	6,5	6,5
	4	-5,5*	5,6	7,8	-3,6*	6,1	7,1	-4,3*	7,2	8,4
	5	-0,6 ns	5,5	5,5	-1,0 ns	5,7	5,8	-0,5 ns	6,3	6,4
180	1	-0,7 ns	5,5	5,6	-0,7 ns	5,0	5,1	-0,5 ns	5,6	5,7
	2	-3,0*	6,2	6,9	-0,4 ns	5,7	5,8	-0,8 ns	7,1	7,1
	3	-1,4 ns	5,3	5,4	-1,3 ns	5,1	5,3	-1,4 ns	6,0	6,2
	4	-5,7*	5,6	8,0	-3,6*	5,7	6,7	-4,1*	7,0	8,1
	5	-0,5 ns	5,4	5,4	-0,8 ns	5,0	5,1	-0,9 ns	5,5	5,6
200	1	-0,2 ns	5,4	5,4	-0,1 ns	5,1	5,1	-0,2 ns	6,0	6,0
	2	-2,5*	5,7	6,2	0,3 s	5,9	5,9	-0,7 ns	7,2	7,2
	3	-0,9 ns	5,3	5,4	-0,7 ns	5,0	5,1	-1,2 ns	6,3	6,4
	4	-5,2*	5,7	7,7	-3,2*	5,5	6,4	-3,1*	7,1	7,7
	5	0 ns	5,4	5,4	-0,2 ns	5,2	5,2	-1,0 ns	6,0	6,1

Note: [a] significance level: * p < 0.05; not significant: ns > 0.05;
Models: 1 – multiplicative; 2 – log(y) ; 3 – sqrt(y); 4 – asin(sqrt(y)); 5 – linear.

Tab.7: Bias, standard error and RMSE for stand- plots estimates of mean volume (m³/ha)

Strip width (m)	Plot area 200 m² bias	std.error	RMSE	400 m² bias	std.error	RMSE	600 m² bias	std.error	RMSE
160	-0,7 ns	13,7	13,7	3,9 ns	14,9	15,4	1,8 ns	18,4	18,4
180	-2,1 ns	14,8	15,0	2,2 ns	14,1	14,3	2,2 ns	17,5	17,6
200	-2,6 ns	14,5	14,8	2,2 ns	14,5	14,7	2,2 ns	18,2	18,4

Note: [a] significance level: * p < 0.05; not significant: ns > 0.05

Regression models comprised two to five predictor variables (Table 8). The most frequently used prediction variable was the interaction term, followed by squared height percentile and canopy density. Generally, the coefficient of determination ranged between 0.79 and 0.96 (Table 9).

Tab.8: Regression models used in each simulation

Strip width (m)	Model	Regression coefficients plot size (m^2) 200	400	600
160	1	bo, b1, b3	bo, b1, b3	bo, b1, b3
	2	bo, b1, b2, b3	bo, b2, b5	bo, b2, b5
	3	bo, b2, b5	bo, b2, b5	bo, b3, b5
	4	bo, b2, b5	bo, b1,b2, b5	bo, b3, b5
	5	bo, b1,b5	bo, b5	bo, b1,b5
180	1	bo, b1, b3	bo, b1, b3	bo, b1, b3
	2	bo, b1, b2, b3, b4	bo, b2, b5	bo, b2, b5
	3	bo, b2, b5	bo, b2, b5	bo, b5
	4	bo, b2, b5	bo, b2, b5	bo, b5
	5	bo, b1, b5	bo, b1, b5	bo, b1,b5
200	1	bo, b1, b3	bo, b1, b3	bo, b1, b3
	2	bo, b1, b2, b3,b4	bo, b2, b5	bo, b1, b2, b5
	3	bo, b2, b5	bo, b2, b5	bo, b2, b5
	4	bo, b2, b5	bo, b2, b5	bo, b2, b5
	5	bo, b1, b5	bo, b1, b5	bo, b5

Regression coefficients: bo - intercept; b1 - canopy density; b2 - (canopy density)^2; b3 - height percentile; b4 - (height percentile)^2; b5 – interaction term (canopy density x height percentile);
Models: 1 – multiplicative; 2 – log(y) ; 3 – sqrt(y); 4 – asin(sqrt(y)); 5 – linear.

Tab. 9: Range of coefficient of determination

Strip width (m)	Model	Plot size (m^2) 200		400		600	
		$R^2_{0,05}$ [a]	$R^2_{0,95}$ [b]	$R^2_{0,05}$ [a]	$R^2_{0,95}$ [b]	$R^2_{0,05}$ [a]	$R^2_{0,95}$ [b]
160	1	0,89	0,95	0,88	0,95	0,88	0,96
	2	0,88	0,94	0,85	0,93	0,86	0,94
	3	0,88	0,94	0,86	0,95	0,88	0,95
	4	0,88	0,94	0,86	0,95	0,88	0,95
	5	0,86	0,94	0,84	0,95	0,85	0,95
180	1	0,88	0,95	0,86	0,96	0,87	0,96
	2	0,88	0,95	0,84	0,93	0,84	0,94
	3	0,89	0,94	0,86	0,95	0,82	0,94
	4	0,89	0,94	0,86	0,95	0,82	0,94
	5	0,87	0,94	0,85	0,95	0,84	0,95
200	1	0,90	0,96	0,87	0,95	0,88	0,96
	2	0,91	0,95	0,86	0,93	0,87	0,96
	3	0,89	0,94	0,88	0,94	0,86	0,95
	4	0,89	0,94	0,88	0,94	0,86	0,95
	5	0,88	0,94	0,87	0,94	0,79	0,94

Note: [a] the 0.05 quantile of empirical distribution; [b] the 0.95 quantile of empirical distribution;
Models: 1 – multiplicative; 2 – log(y) ; 3 – sqrt(y); 4 – asin(sqrt(y)); 5 – linear.

The bias of mean volume estimates during iterations in each simulation ranged between -16.6 m^3/ha (6.5%) and 10.2 m^3/ha (4.0%) for regression estimates (Table 10), while the bias of ground-based estimates (Table 11) ranged from -34.1 m^3/ha (13.4%) to 31.8 m^3/ha (12.5%).

The MC estimates for mean volume derived by regression ranged between -5.7 m^3 (2.2%) and 0.3 m^3/ha (0.1%), and standard error between 5.0 m^3/ha (2.0%) and 7.4 m^3/ha (2.9%). For plot-based MC estimates, the range of bias was between -2.6 m^3/ha (1.0%) and 3.9 m^3/ha (1.5%), with a standard error between 13.7 m^3/ha (5.4%) and 18.4 m^3/ha (7.2%). The lowest RMSE for regression-based MC estimates was 5.1 m^3/ha (2.0%) and the highest was 8.4 m^3/ha (3.3%). RMSE for ground-plot MC estimates varied between 13.7 m^3/ha (5.4%) and 18.4 m^3/ha (7.2%).

Tab. 10: Range of bias for regression estimates

Model	Stripe bredde	Plot size (m^2) 200				400				600			
		bias $_{0,05}$ [a]		bias $_{0,95}$ [b]		bias $_{0,05}$ [a]		bias $_{0,95}$ [b]		bias $_{0,05}$ [a]		bias $_{0,95}$ [b]	
	(m)	m^3/ha	%	m^3/ha	%	m^3/ha	%	m^3/ha	%	m^3/ha	%	m^3/ha	%
1		-11,4	-4,5	5,7	2,24	-11,3	-4,4	9,4	3,7	-10,5	-4,1	9,3	3,7
2		-16,1	-6,3	5,4	2,13	-12,3	-4,8	8,4	3,3	-13,6	-5,4	10,2	4
3	160	-10,1	-4	5,6	2,2	-11	-4,3	8,1	3,2	-12,5	-4,9	8,7	3,4
4		-15,9	-6,3	2,6	1,02	-14,2	-5,6	7	2,8	-16,4	-6,5	6,2	2,4
5		-11	-4,3	7,7	3,03	-9,9	-3,9	9,4	3,7	-11,6	4,6	9,1	3,6
1		-10,6	-4,2	7,8	3,07	-8,2	-3,2	7,9	3,1	-8,9	-3,5	7,9	3,1
2		-12,9	-5,1	6,5	2,56	-8,7	-3,4	9,3	3,7	-12,7	-5	9,6	3,8
3	180	-10,4	-4,1	7,4	2,91	-8	-3,1	7,8	3,1	-11,5	-4,5	7,3	2,9
4		-15,5	-6,1	1,4	0,55	-12,4	-4,9	8,7	3,4	-14,3	-5,6	6	2,4
5		-10,4	-4,1	9,1	3,58	-7,8	-3,1	8,2	3,2	-9,1	-3,6	7,1	2,8
1		-10,3	-4,1	8,6	3,39	-8,1	-3,2	8,7	3,4	-11	-4,3	9	3,5
2		-12,4	-4,9	6,7	2,64	-9,4	-3,7	9,7	3,8	-13,3	-5,2	9,2	3,6
3	200	-11,2	-4,4	6,3	2,48	-7,8	-3,1	8,4	3,3	-11,9	-4,7	8	3,1
4		-16,6	-6,5	3,1	1,22	-12,1	-4,8	7,8	3,1	-15,1	-5,9	7,7	3
5		-11,3	-4,4	7,3	2,87	-7,4	-2,9	8,6	3,4	-10,8	-4,3	7,9	3,1

Note: [a] the 0.05 quantile of empirical distribution; [b] the 0.95 quantile of empirical distribution;
Models: 1 – multiplicative; 2 – log(y) ; 3 – sqrt(y); 4 – asin(sqrt(y)); 5 – linear.

Tab. 11: Range of ground- based systematic plot sampling derived estimates

Strip	Plot size (m^2) 200				400				600			
width	bias $_{0,05}$ [a]		bias $_{0,95}$ [b]		bias $_{0,05}$ [a]		bias $_{0,95}$ [b]		bias $_{0,05}$ [a]		bias $_{0,95}$ [b]	
(m)	m^3/ha	%	m^3/ha	%	m^3/ha	%	m^3/ha	%	m^3/ha	%	m^3/ha	%
160	-23,7	-9,3	22,8	9	-20,7	8,1	26,7	10,5	-26	-10,2	28,2	11,1
180	-34,1	-13,4	17,9	7	-23,4	-9,2	24,1	9,5	-18,9	-7,4	29,8	11,7
200	-27,6	-10,9	19,4	7,6	-23,4	-9,2	22,1	8,7	-22,6	-8,9	31,8	12,5

Note: [a] the 0.05 quantile of empirical distribution; [b] the 0.95 quantile of empirical distribution;

Among all regression models, only the multiplicative and linear models gave unbiased estimates ($p > 0.05$). Ground-based systematic plot sampling derived estimates provided unbiased estimates ($p > 0.05$) for all sampling designs.

Relative efficiency of laser-based estimates relative to ground-plot estimates varied between 0.33 and 0.57 (Table 12).

Tab.12: Relative efficiency (RMSE[a] / RMSE[b]) of laser- based against ground- plots estimates

Strip width (m)	Model	Plot size (m^2) 200	400	600
	1	0,41	0,38	0,35
	2	0,48	0,42	0,40
160	3	0,39	0,38	0,35
	4	0,57	0,46	0,46
	5	0,40	0,38	0,35
	1	0,37	0,36	0,32
	2	0,46	0,41	0,40
180	3	0,36	0,37	0,35
	4	0,53	0,47	0,46
	5	0,36	0,36	0,32
	1	0,36	0,35	0,33
	2	0,42	0,40	0,39
200	3	0,36	0,35	0,35
	4	0,52	0,44	0,42
	5	0,36	0,35	0,33

Note: [a] RMSE of laser- based estimates; [b] RMSE of laser- based estimates
Models: 1 – multiplicative; 2 – log(y) ; 3 – sqrt(y); 4 – asin(sqrt(y)); 5 – linear.

5. Discussion

The results indicated that forest surveys over large areas carried out using airborne laser scanning as a strip sampling tool can provide accurate estimates, and can be more effective than traditional systematic ground-plot based inventories.

An important factor which influenced the simulation results was the forest stand structure, reflected by the (x_i, y_i)-locations of each tree within study area. The software used to generate the virtual forest area(the program package SILVA2.2) is currently used for forest stand growth simulation in Germany(Pretzsch et al., 2002). The single tree-based stand simulator included in this package has been developed and parameterised using more than 155,000 tree observations from forest research and trial plot inventory data in Germany, and it was found to be most reliable particularly for site conditions found in southern Germany(Pretzsch et al., 2002).

No evaluations of SILVA2.2 under Nordic climate and site conditions are known. Anyway, an assessment of the reliability of this software was previously done under Swiss conditions . According to this research, SILVA2.2 performs adequately from the hills to the mountain zone, and supports extrapolations even for elevations above 1000 m in areas with no extreme climate, and therefore it was considered to be appropriate for the scope of this project.

For a given forest structure, the simulation results varied with sampling design parameters, which had rather a combined influence against mean volume estimates.

First, the strip location during a given sequence of iterations is influenced by the strip width, because the sampling procedure is designed in such a way that the sampling strips will never cross the borders of study area. This implies that under each simulation using sampling schemes with the same strip width, the sampled strips will be positioned at the same locations for all iteration sequences, but locations will differ when the strip width changes and therefore other trees will be sampled. Thus, ground-based estimates varied for schemes using same strip width, but these variations are not functionally correlated to the size of strip width.

Second, there is an inverse relationship between ground plot size and number of plots in each sample, since the intention was to hold the field sampling intensity at same levels for all schemes. As the size of stand plots increases, the sample size decreases and then the conventional estimates and regression coefficients would probably be less precise. This effect can be compensated by the ground plots' capacity to reflect the variability of the forest structure and to reduce between-plot variability. As plot size increases, within-plot variability increases and between-plot variability decreases, resulting in a smaller variance estimate across all plots. The estimation method only accounts for the between-plot variance, and then the ability of plot size to reflect the within-variance will influence the precision of the volume estimate. Furthermore, when smaller samples sizes comprising larger plots are used to build regression equations, the number of cells used for prediction will decrease as well, because the ground-based systematic sampling plots and grid cells had the same area. Therefore, increasing strip width tends to improve the predictions, especially for sampling schemes using larger ground sampling plots.

Since the purpose of regression models was estimation and prediction and not interpretation of cause-effect relationships, all stepwise models were considered appropriate for prediction, even though they were susceptible of serious multicollinearity problems. Arguably, stepwise procedure performed worse than best subsets method. The results from Table 5 cannot be generalised over all simulation results. However, they showed that by employing stepwise regression in presence of multicollinearity, there is a chance of selecting slightly biased subsets. In the present study, the bias was not significant (Table 6), but this might be a consequence of using a very uniform dataset. Indeed, the study area comprised more than 2.7 million trees, which were derived from only 435 single-tree models, and only 38 forest stand models of 1.0 ha each were used to represent a study area of 3600 ha. For sampling of populations with large variation of individuals, stratification will probably improve the performance of stepwise regression and reduce the amount of bias.

Because the multiplicative model was not subject to stepwise selection procedure, the performance of this model was assessed against the ground-based systematic plot sampling method, for all sampling schemes. However, during exploratory trials, the multiplicative model yield highest adjusted-R^2 and prediction-R^2 in comparison to the other models for samples with plot size of 200 m^2 and 600 m^2, and average adjusted-R^2 and prediction-R^2 values for plots of 400 m^2. For the reasons mentioned above, the choice of multiplicative models seemed to be a reliable alternative to other regression models, under the assumption that the model is unbiased after the bias correction is applied for back-transformation to arithmetic scale.

Both inventory methods, i.e. ground-based systematic plot sampling and laser scanning-based strip sampling employing multiplicative regression model, provided unbiased MC-estimates ($p > 0.05$) of mean volume. The regression model performed clearly better than the ground-based systematic plot sampling. The RMSE for regression-based MC estimates ranged from 5.1 m^3/ha (2.0%) up to 6.4 m^3/ha (2.5%), comparing to a range from 13.7 m^3/ha (5.4%) to 18.4 (7.2%) m^3/ha for plot-based MC estimates. The regression-based MC estimates had generally slightly underestimated the true mean volume.

Considering the size of the population, the bias and precision of the MC estimates of mean volume has shown little variation for different sampling designs (Appendix D-Figure 5). The MC estimates seemed to be robust against variation of plot sample size and grid cell area. Regression-based

MC estimates varied within 0.8 m^3/ha from the population mean, while the ground-based systematic plot sampling estimates varied within 3.9 m^3/ha from the true value.

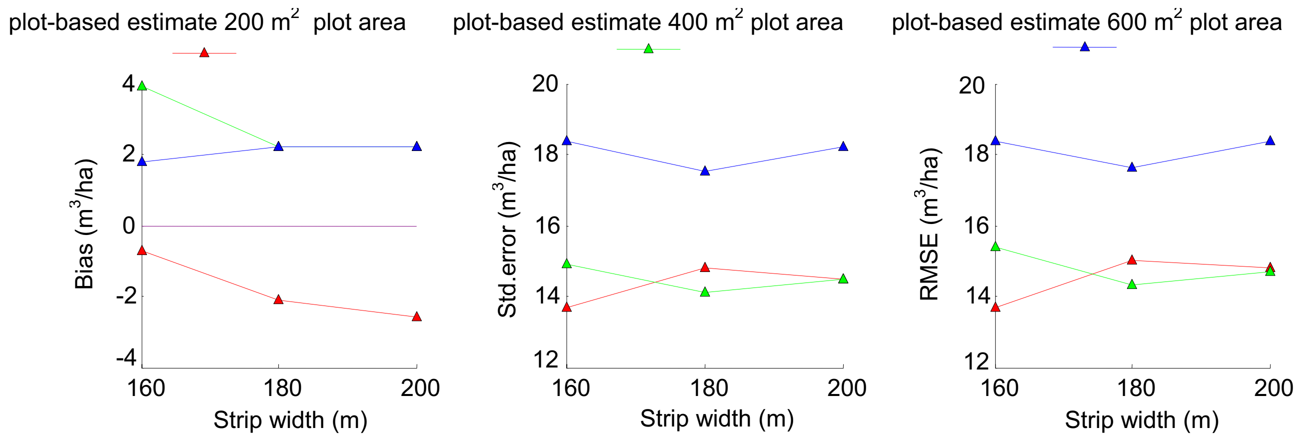


Fig.5: Variation of ground-based systematic plot sampling MC-estimates with strip width

For the ground-based systematic plot sampling method, the plot size was the dominant factor which led the overall trends for the mean volume MC-estimates (Figure 5). By increasing strip width, small fluctuations for bias and standard error occurred for similar plot sizes, because the MC-estimates were based on different samples.

A better differentiation of sampling designs occurred when the plot size was increased. MC-estimates derived from ground-based systematic sampling with plot sizes of 200 and 400 m^2 had the lowest RMSE values-between 13.7 and 15.4 m^3/ha, while for plots of 600 m^2 the RMSE values increased to between 17.6 and 18.4 m^3/ha because of plot number reduction. The MC ground-based estimates obtained using 200 m^2 plots were underestimating the population mean volume with values from 0.7 to 2.6 m^3/ha. The ground-based systematic sampling schemes using plot sizes of 400 and 600 m^2 had relative similar fluctuations of bias, with a maximum of 3.9 m^3/ha registered for plots of 400 m^2. The largest RMSE values (18.4 m^3/ha) were noticed for plots of 600 m^2, probably as a consequence of smaller number of plots in each sample.

By increasing the sample size with almost 300%, from 36 plots of 600 m^2 to 102 plots of 200 m^2, the improvement in precision for MC ground-based estimates was around 25%, and it did not provide a more accurate mean volume estimate. Furthermore, no significant loss in precision was registered when the sample size was reduced from 102 plots of 200 m^2 to 51 plots of 400 m^2.

For regression based MC-estimates, the variation of standard deviation and RMSE followed the curve shapes as for conventional ground-based samples, but the differences between outcomes from each sampling scheme were smoothed (Figure 6). The RMSE for all estimates differed with less that 1.3 m^3/ha, and the bias did not vary with more than 0.7 m^3/ha.

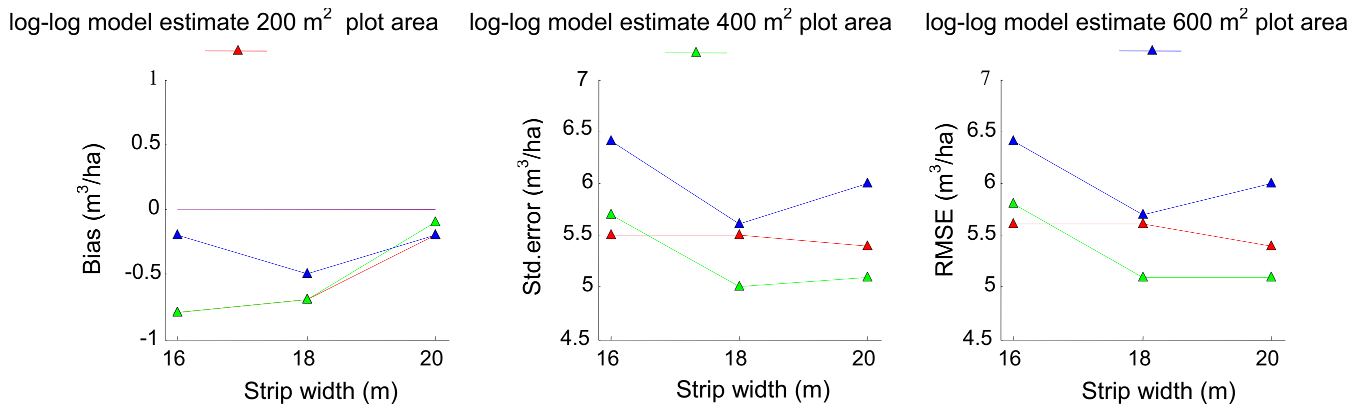


Fig.6: Variation of regression – based MC estimates of mean volume with strip width

The accuracy of regression based MC-estimates generally increased with strip width, probably due to a larger number of grid cells for prediction of stripe volume. The accuracy of the MC-estimates based on schemes using plots of 200 and 400 m^2 increased with strip width. The most significant change in bias occurred when increasing the strip width from 180 to 200 m, when the bias reduction was from -0.7 to -0.1 m^3/ha for sampling schemes with plots of 400 m^2, from -0.7 to -0.2 m^3/ha for sampling schemes with plots of 200 m^2 and from -0.5 to -0.2 m^3/ha for sampling schemes with plots of 600 m^2. Increasing the strip width from 160 to 180 had the highest effect on bias for the MC-estimates based on plots of 600 m^2, which increased from -0.2 to -0.5 m^3/ha. Significant changes in the level of precision was noticed when the strip width increased from 160 to 180 m, for sampling schemes using plots of 400 and 600 m^2, where RMSE decreased from 5.8 to 5.1 m^3/ha and from 6.4 to 5.7 m^3 respectively.

Finally, the relative efficiency of regression-based laser scanning estimates relative to the correspondent ground-based estimates (Table 12) increased for all regression models, when strip width and plot size increased. In this analysis, laser-derived MC-estimates were from 1.8 and up to 3.0 times more efficient than their correspondent ground-based estimates. The main cause of this effect was probably that the RMSE of the ground-based estimates was degrading more rapidly than the RMSE of the laser-based MC estimates, when the plot size increased. Since the field inventory for this study had a much higher intensity comparing to operational surveys, the efficiency of ground-based estimates may become further degraded in the situation of a real scale forest sampling, when the number of ground plots will become object to economical constrains.

In conclusion, the present study has indicated that the laser scanning-based strip sampling method was suitable for volume assessment of a theoretical model forest. Moreover, this method performed better the ground-based systematic plot sampling. The laser scanning based method using multiplicative regression models achieved double precision of mean volume estimate, and the estimates were unbiased. Strip sampling schemes using small plots generally gave accurate MC-estimates, which seemed not to be influenced by strip width variations. The influence of the strip width became more significant for large plots, when the precision of MC-estimates generally tends to increase as the strips become wider and strip sampling intensity increases.

Laser based estimates had in average a double relative efficiency comparing to traditional sampling procedure. However, generalizations cannot be drawn from this study, since many assumptions were not realistic compared to real-world applications, i.e. size of the target area, variability of the model population, and sampling intensity. Another important issue is that all metrics derived from the population were considered to be “error free” and the effects of error propagation were not neglected. As possible error sources could be mentioned errors concerning ground location of trees and ground plots, laser sampling and field measurements.

To improve simulator’s performances, it should be developed a forest stand generator calibrated on Norwegian site particularities. The simulator should also be able to handle large forest areas (i.e. at least 100 km^2). Other major challenges might be to develop laser scanning-based inventory procedures adapted for other tree species than Norway spruce, and to a large empirical database of laser-derived individual tree models for all main tree species in Scandinavia.

Acknowledgments

I would like to thank my supervisors, Professor Erik Næsset and Assistant Professor Terje Gobakken, Univ. of Life sciences, for the constant support and huge amount of patience they have shown to me during the work with this project.

Thank you!

Liviu Theodor Ene,
Ås, 30.April, 2007

References

"Matlab(R2006a) documentation." Retrieved 02.feb, 2007, from www.mathworks.com.

Bollandsås, O. M., Næsset,E. (2007). "Estimating percentile-based diameter distributions in uneven-sized Norway spruce stands using airborne laser scanner data." Scandinavian Journal of Forest Research **22**:33-47 .

Gobbaken, T., Næsset, E, Nelson, R (2006). "Developing regional forest inventory procedures based on scanning LiDAR." SilviLaser Meeting, Matsuyama, Ehime (Japan).

Magnussen, S., Boudewyn, P. (1998). "Derivations of stand heights from airborne laser scanner data with canopy-based quantile estimators." Canadian Journal of Forest Research **28**: 1016-1031

Mendenhall, W., Sincich, T. (2003). A second course in statistics. Regression analysis, Pearson Education, Inc., p.329

Nelson, R., Short, A., Valenti, M. (2004). "Measuring biomass and carbon in Delaware using an airborne profiling LIDAR." Scandinavian Journal of Forest Research **19**: 500-511.

Nelson, R., Næsset, E., Gobakken, T., Ståhl, G., Gregoire T. G. (2006). "Regional forest inventory using an airborne profiling LiDAR." SilviLaser Meeting, Matsuyama, Ehime (Japan).

Næsset, E. (1997). "Determination of mean tree height of forest stands using airborne laser scanner data." ISPRS Journal of Photogrammetric Engineering & Remote Sensing **52**: 49-56.

Næsset, E. (2002). "Predicting forest stand characteristics with airborne scanning using a practical two- stage procedure and fiels data." Remote Sensing and Environment **80**: 88-199.

Næsset, E., Bjerknes, K. (2001). "Estimating tree heights and number of stems in young forest stands using airborne laser scanner data." Remote Sensing and Environment **78**: 328-340.

Næsset, E. (2004). "Practical large-scale forest stand inventory using small-footprint airborne scanning laser." Scandinavian Journal of Forest Research **19**: 164-179.

Næsset, E., Gobbaken, T., Holmgren, J., Hyyppä, H., Hyyppä, J., Maltamo, M., Nilsson, M., Olsson, H., Persson, Å., Söderman, U. (2004). "Laser scanning of forest resources: The Nordic experience." Scandinavian Journal of Forest Research **19**: 1-18.

Pretzsch, H., Biber, P., Ďursky, J. (2002). "The single tree-based stand simulator SILVA: construction, application and evaluation." Forest Ecology and Management **162**: 3-21.

Schmidt, S., Zingg, A., Biber, P. (2006). "Evaluation of the forest growth model SILVA along an elevational gradient in Switzerland." European Journal of Forest Research **125**: 43-55.

Solberg, S., Næsset, E., Bollandsås, O.M. (2006). "Single tree segmentation using airborne laser scanner data in a structurally heterogenous spruce forest." ISPRS Journal of Photogrammetric Engineering & Remote Sensing **72**(12): 1369-1378.

Wynne, R. H. (2006). "Lidar remote sensing of forest resources at the scale of management" ISPRS Journal of Photogrammetric Engineering & Remote Sensing **72**(12): 1311-1314.

Vestjordet, E. (1967). "Functions and tables for volume of standing trees. Norway spruce." Communications of Norwegian Forest Research Institute **22**: 543-574. (In Norwegian with English summary.)

Appendix A

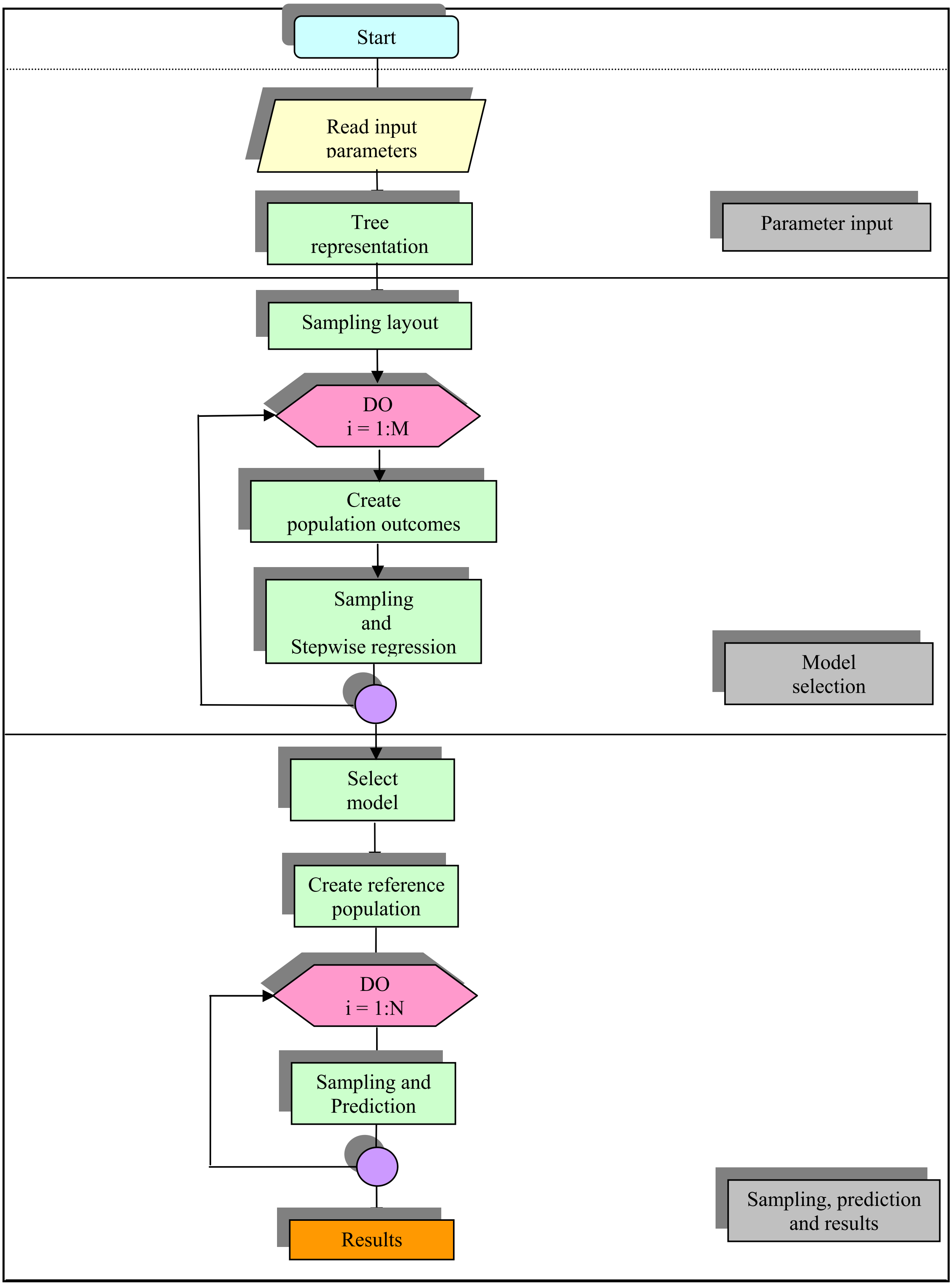
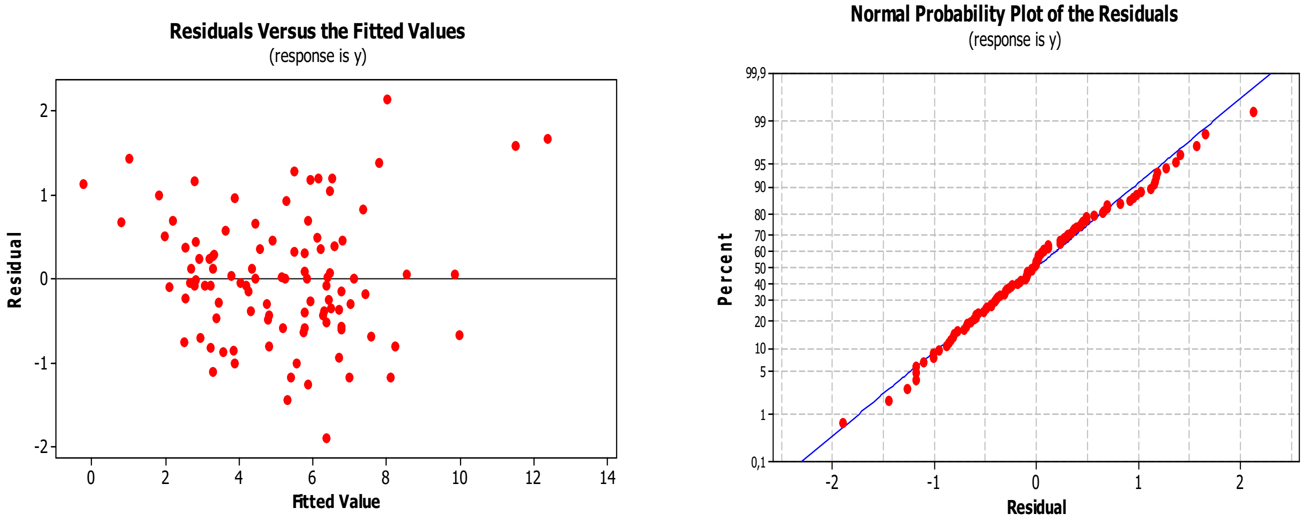
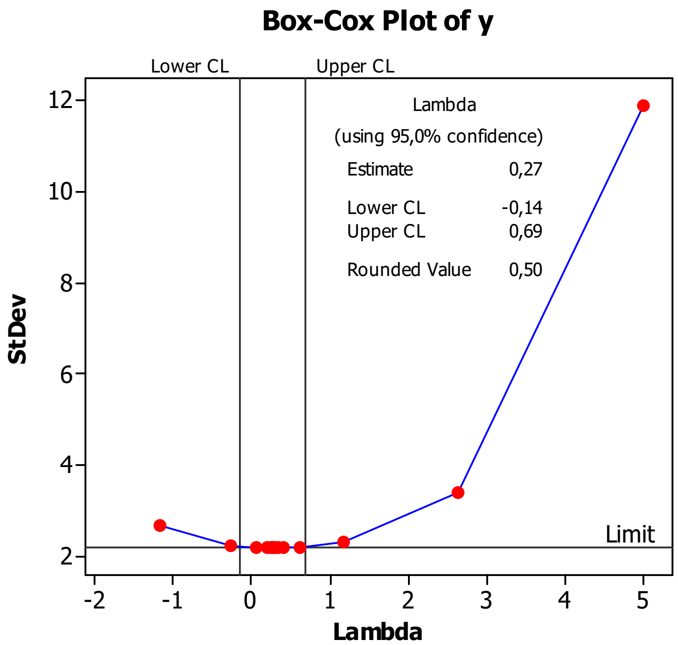


Fig.3: Flow chart for sampling simulation

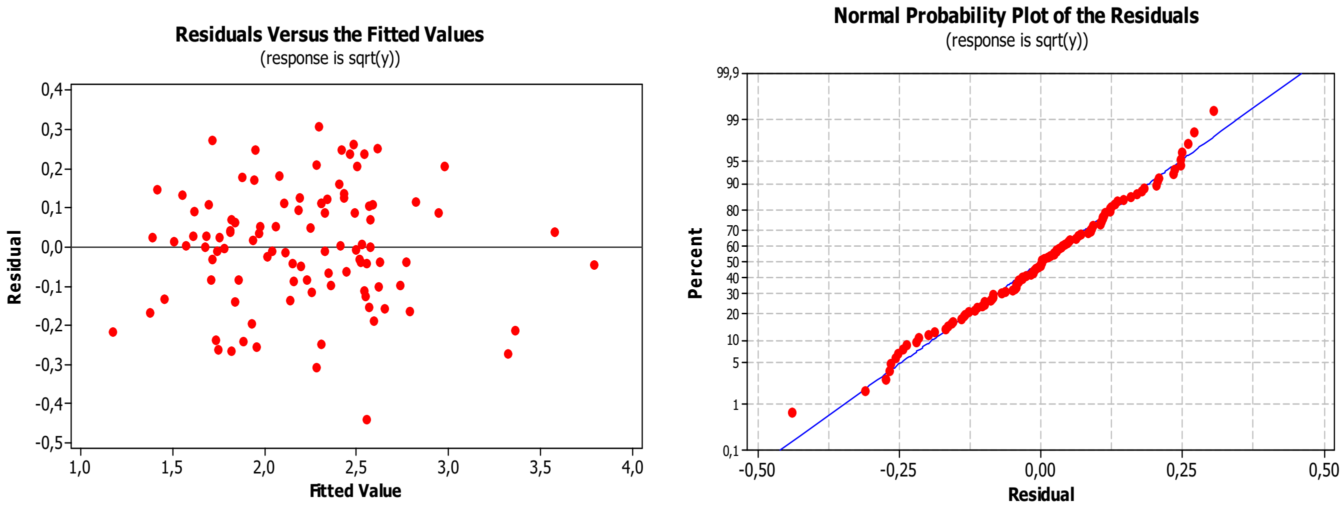
Appendix B



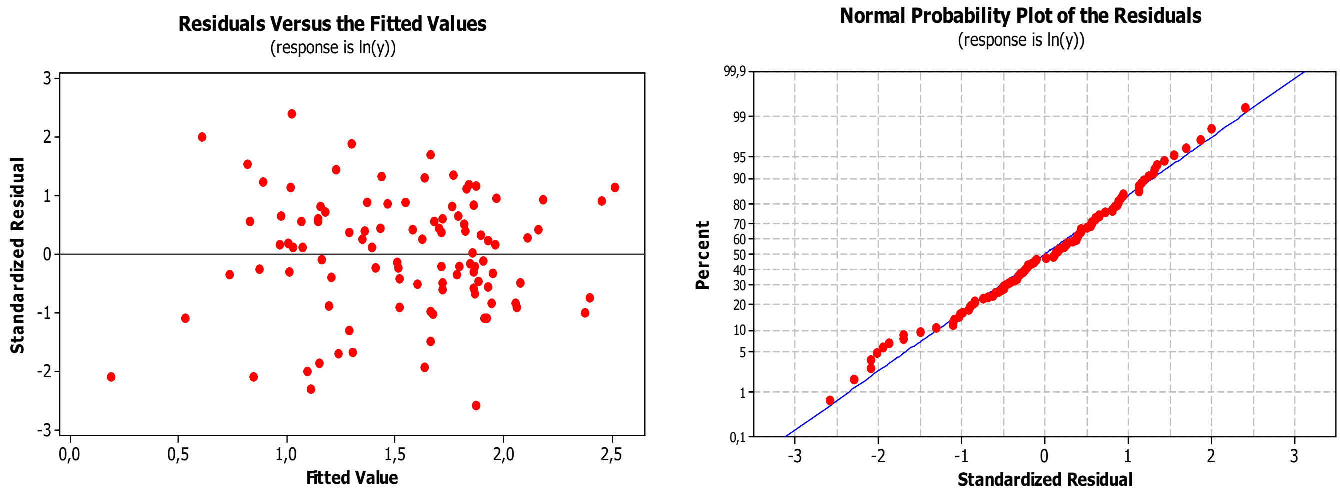
a.



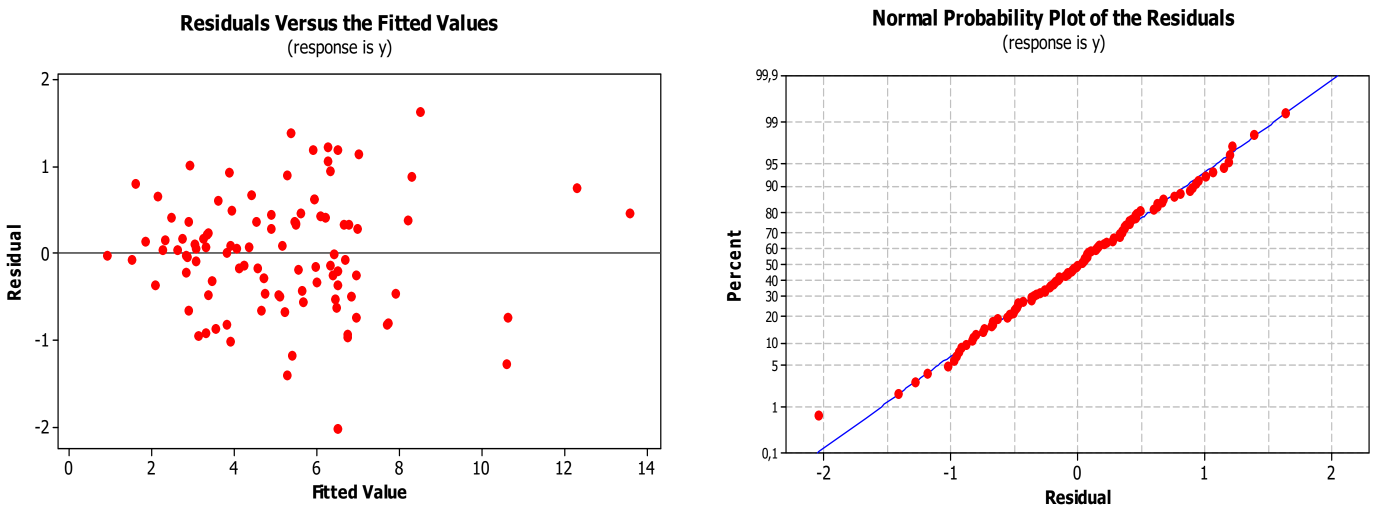
b.


Residuals Versus the Fitted Values
(response is sqrt(y))
Residual
Fitted Value
Normal Probability Plot of the Residuals
(response is sqrt(y))
Percent
Residual

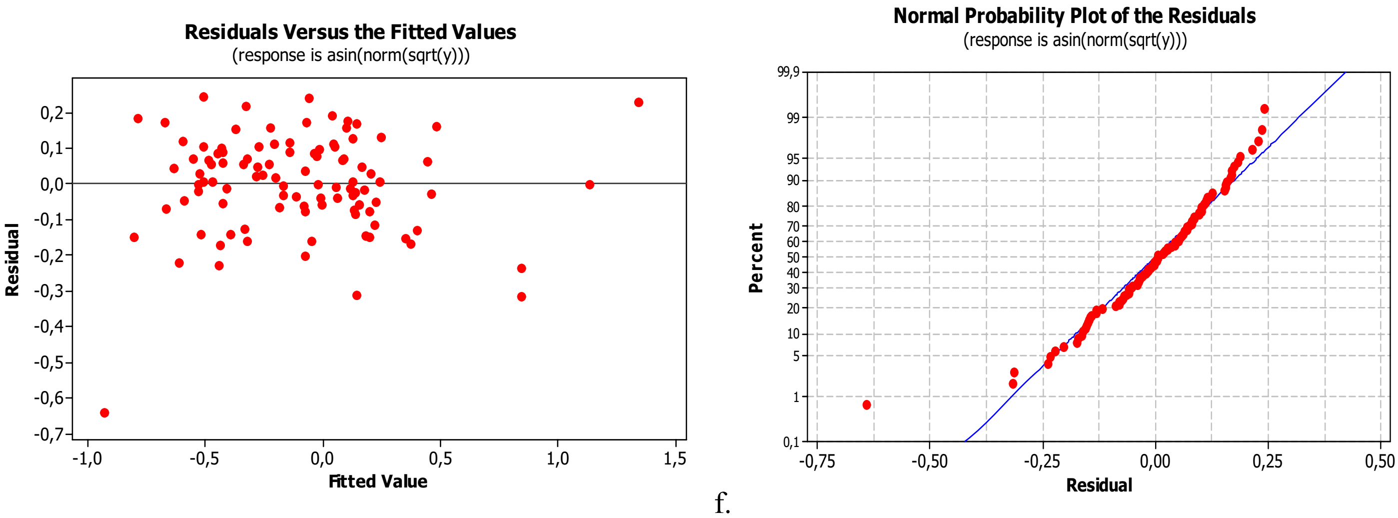
c.


Residuals Versus the Fitted Values
(response is ln(y))
Standardized Residual
Fitted Value
Normal Probability Plot of the Residuals
(response is ln(y))
Percent
Standardized Residual

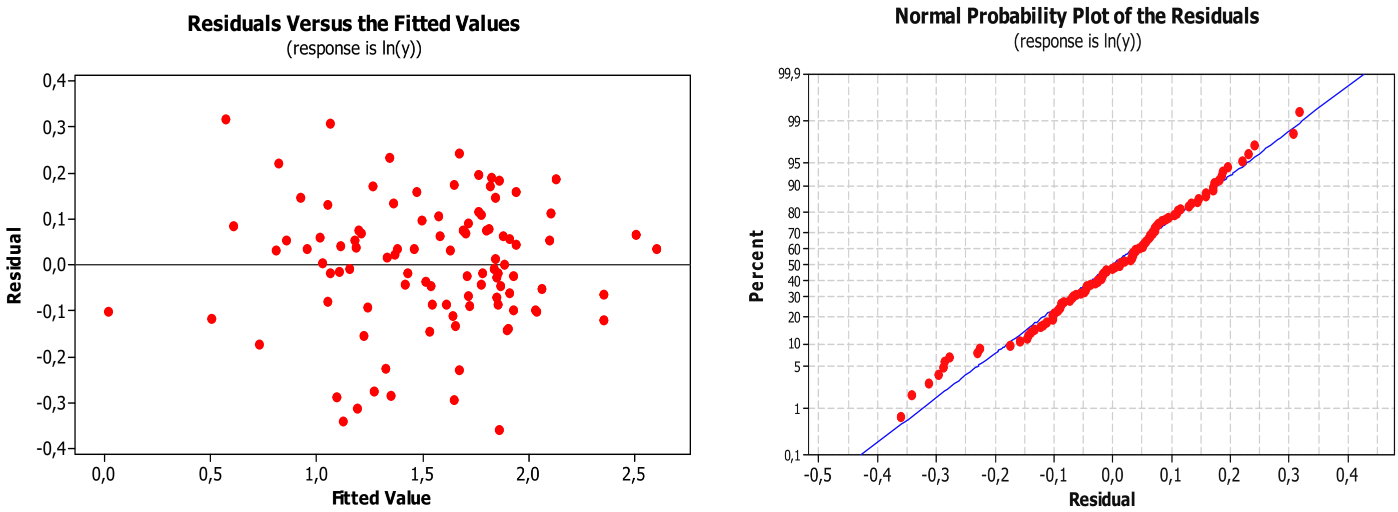
d.


Residuals Versus the Fitted Values
(response is y)
Residual
Fitted Value
Normal Probability Plot of the Residuals
(response is y)
Percent
Residual

e.



f.



g.

Fig. 4 : Residual and normal plots for main effects model (a);

Boc-Cox transformation (b);

Residual and normal plots for stepwise regression subset models:

- (c) linear;
- (d) sqrt(y);
- (e) log(y);
- (f) asin (sqrt(y));
- (g) multiplicative.

Appendix C

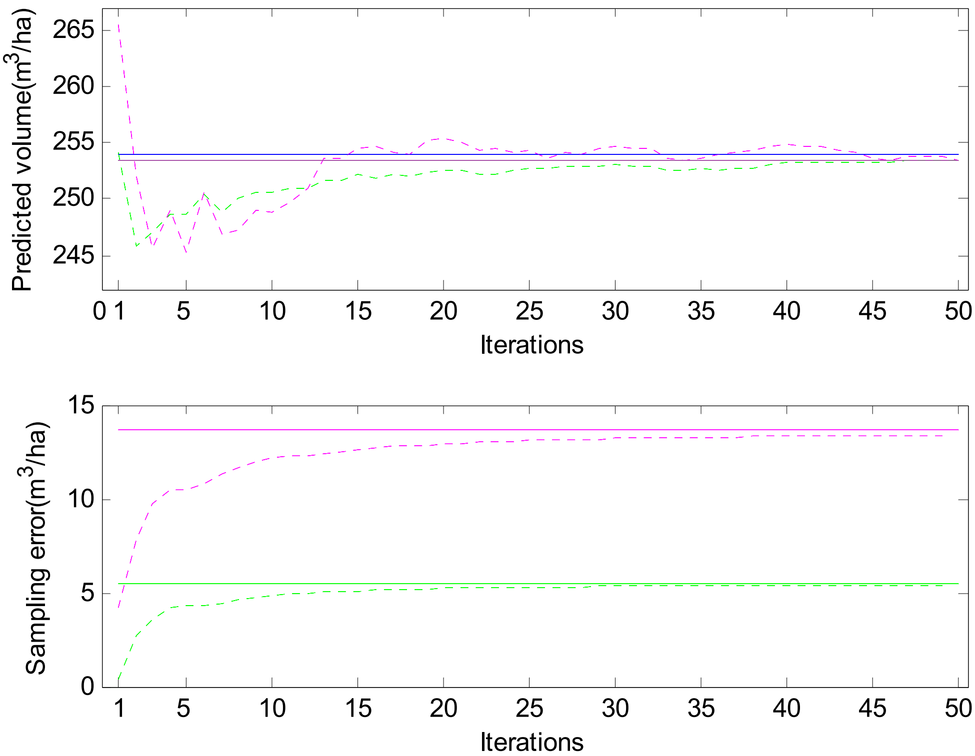
Monte Carlo estimation of population mean and sampling error, for multiplicative regression model and ground plot base estimates:

Legend:

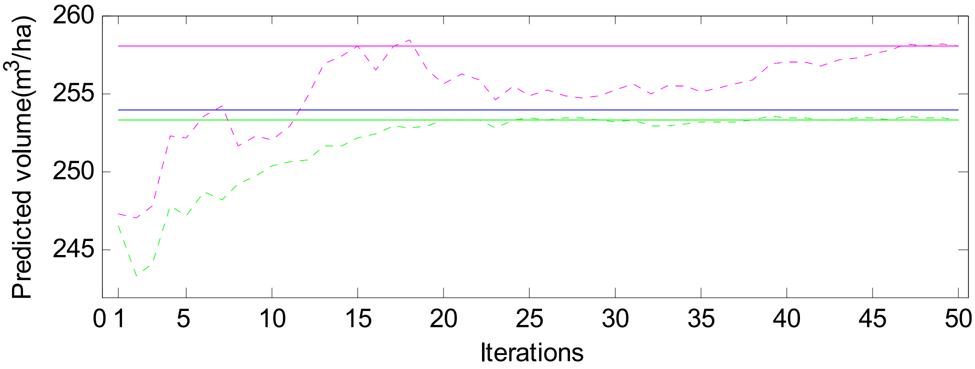
Cumulated regression based estimate over iterations
Cumulated ground plot based estimate over iterations
Regression based Monte Carlo estimate after 50 iterations
Ground plot based Monte Carlo estimate after 50 iterations
Population mean

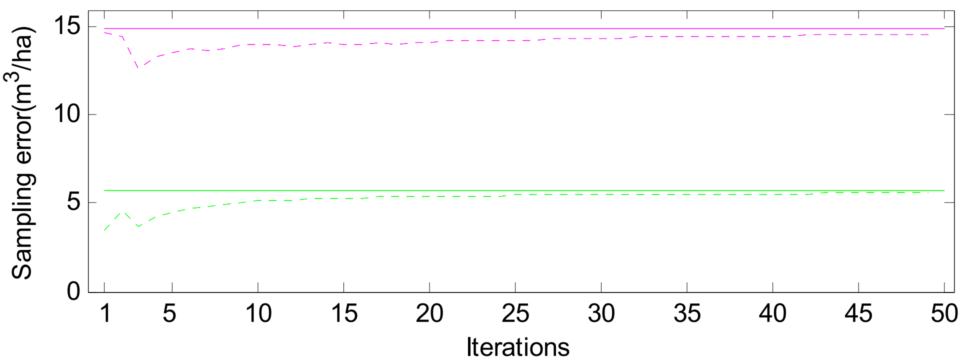
1. strip width = 160 m;

   a) plot area = 200 m^2

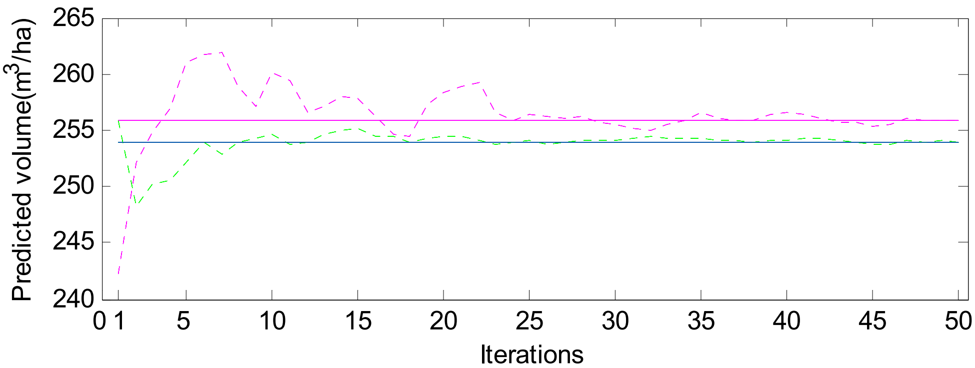


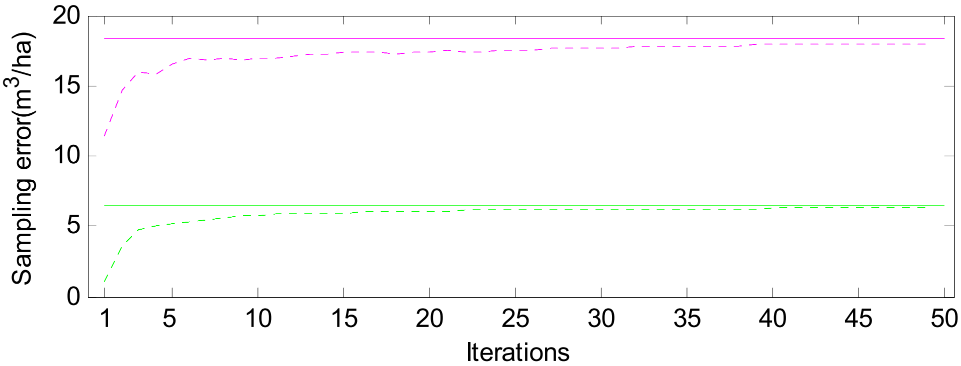
b) plot area = 400 m^2





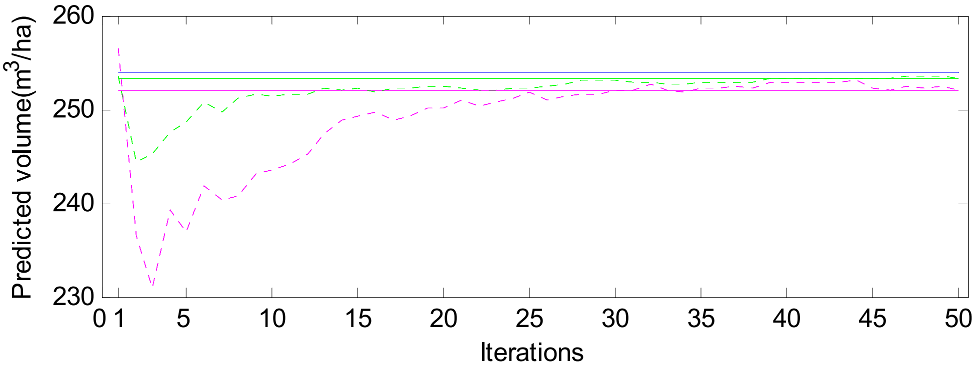
c) plot area = 600 m^2

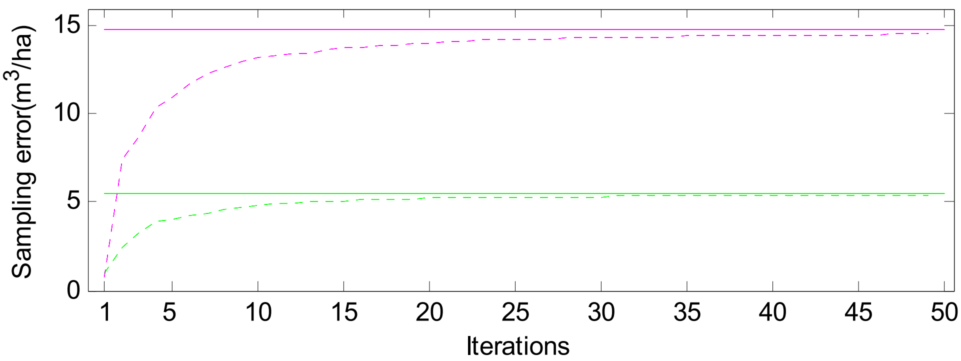




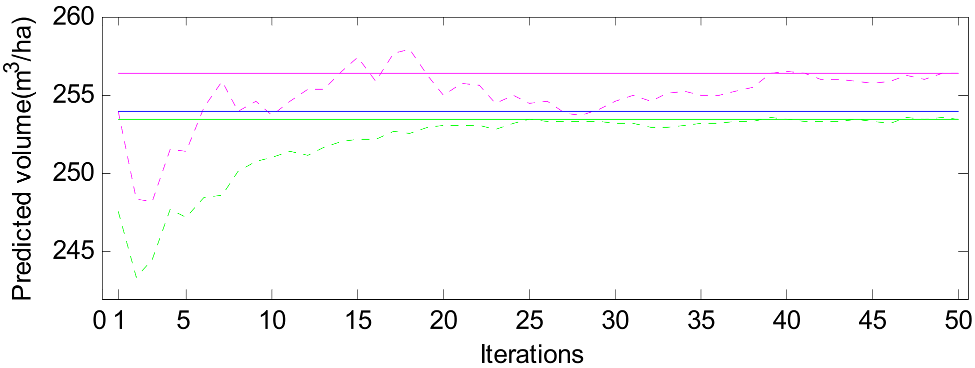
2. strip width = 180 m;

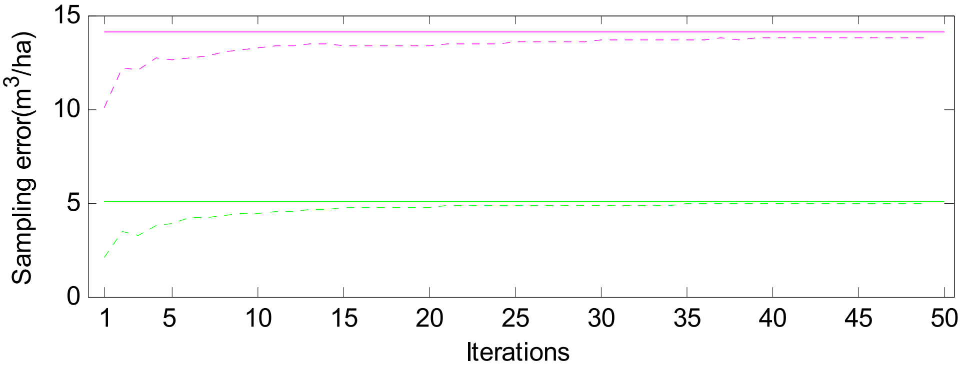
a) plot area = 200 m^2;



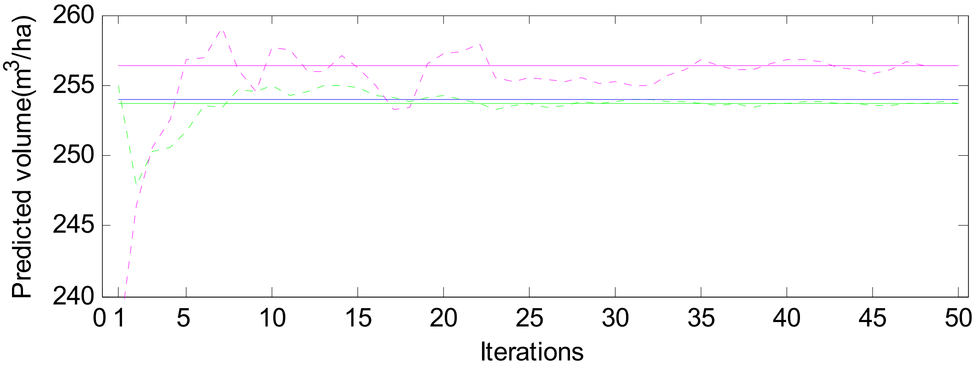


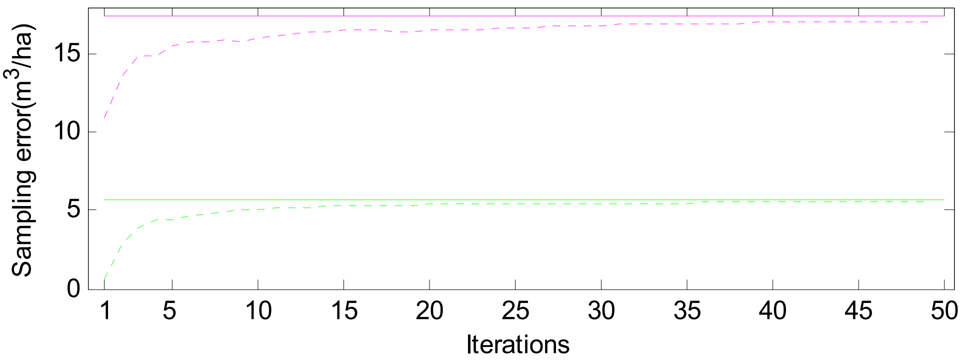
b) plot area = 400 m^2;





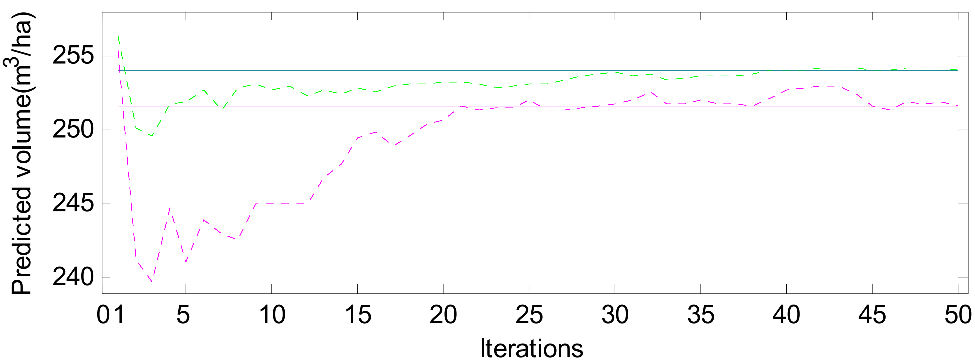
c) plot area = 600 m^2;

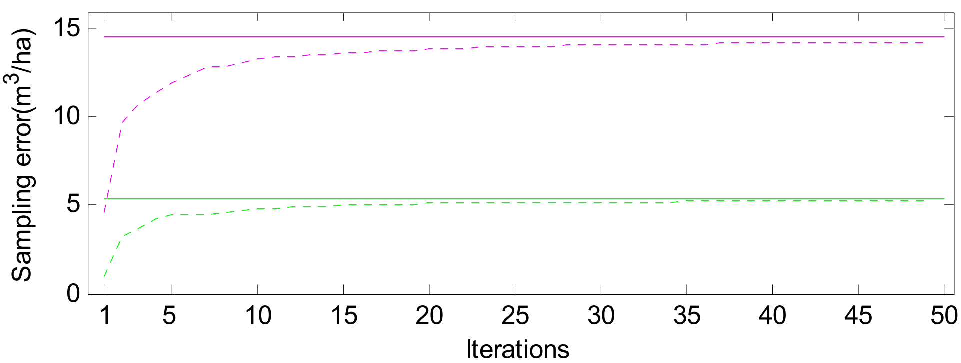




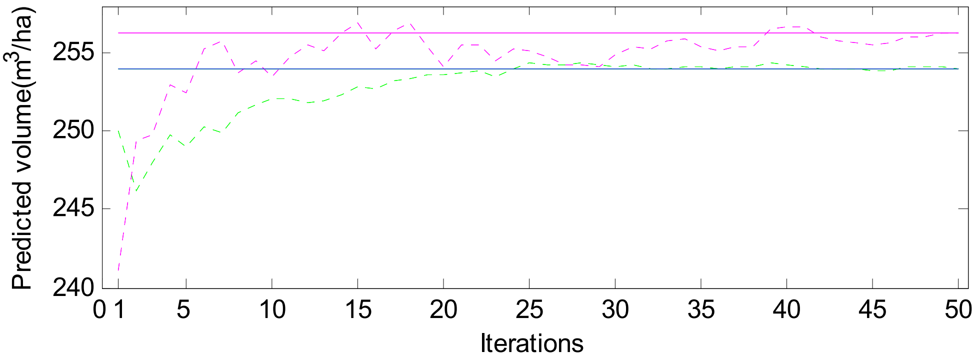
3. strip width = 200 m;

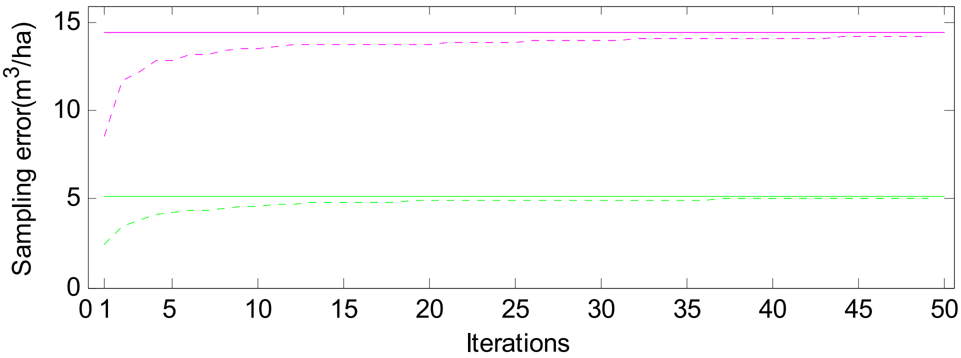
a) plot area = 200 m^2;



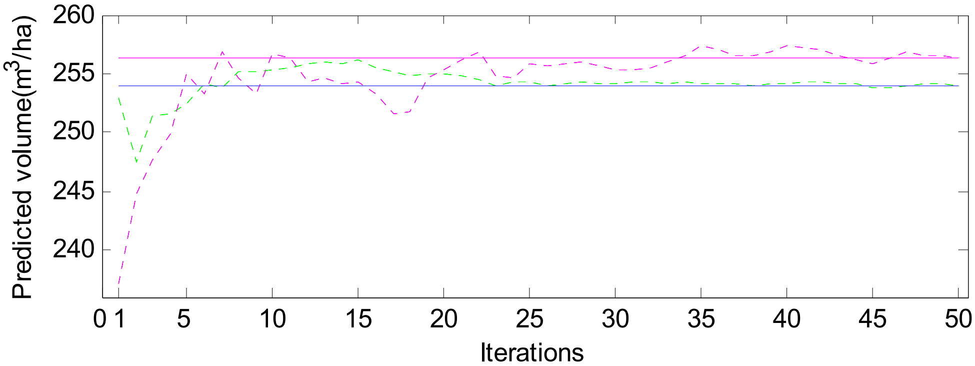


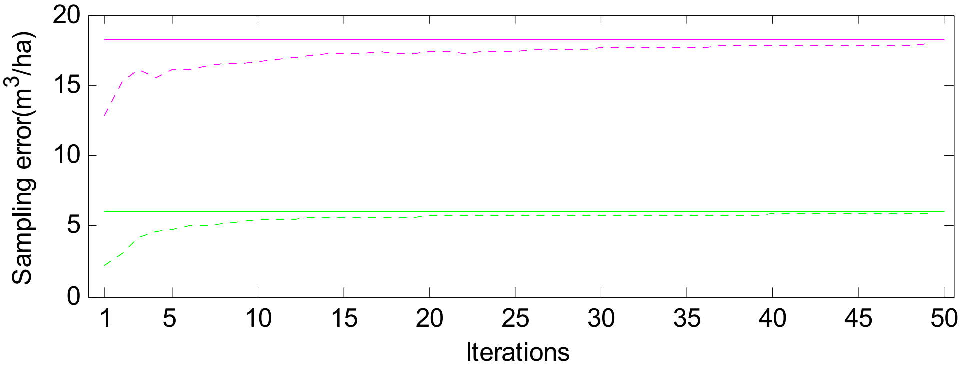
b) plot area = 400 m^2;





c) plot area = 600 m^2;





Appendix D

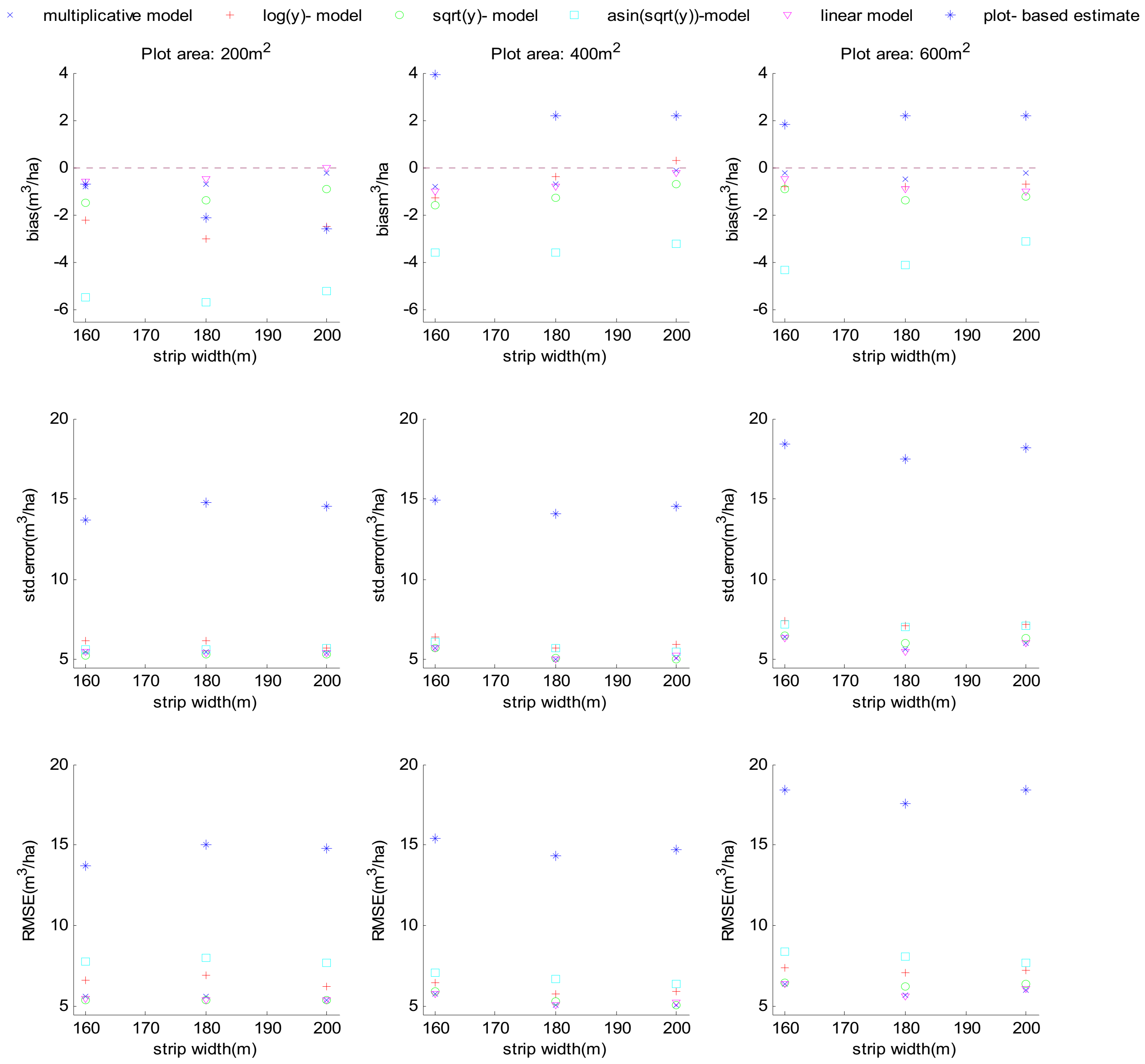


Fig.5: Bias, standard error and RMSE for regression and ground-plots MC mean volume estimates