Quality Control of Measurements Made on Fixed-Area Sample Plots

Ola Lindgren

Abstract.—The paper describes results from a large program for quality control of forest measurements. The performance of 87 surveyors was evaluated. Tree heights were usually measured well, whereas the counting of tree-rings on increment cores was a source of considerable bias for many surveyors. During tree count on sample plots, many surveyors had a tendency to forget trees, especially small trees. A conclusion of the study is that control surveys are justified. The control should be continuous during the field season, making it possible to detect inferior quality of work at an early stage.

This study was done to summarize the findings of control surveys that are regularly carried out and analyzed by our company. The forest surveys being checked were applications of field data collection for the Swedish "Forest Management Planning Package" (Jonsson *et al.* 1993). The original field data were collected from large forest holdings by a stratified multi-stage sampling design, using delineated stands as primary sampling units. Within each sample stand, 6 to 12 circular plots were laid out in a square grid pattern.

A survey of this kind typically estimates the growing stock with a sampling error of less than 3 percent at the forest holding level. It is therefore of interest to also estimate measurement errors, since they might have the same effect on the final estimates as the sampling errors. In addition, measurement errors don't always have an expectation of zero, which is the case for the sampling errors when an adequate probability sampling plan is applied.

By summarizing the performed control surveys, we are looking for answers to several questions. What are the characteristics of the measurement errors? How do different field surveyors compare in terms of quality of work? Are control surveys justified, or just a waste of money? If justified, how do we carry out the control program in the most efficient way?

THE DATA SET

The data set was derived from control surveys made during 1991-1997 within 10 different field projects. A total of 87 field surveyors were checked. The data consist of 2,523 fixed-area circular sample plots of radius 5-10 m, in which a total of 47,457 trees were measured by caliper. Several 3- to 10-m control plots were located in each of 529 randomly selected stands. Measurements of total tree height and age at dbh were controlled for 3,356 subsample trees. A small study of spatial accuracy in plot location was made on 46 sample plots.

The surveys were made in central and northern Sweden in forests dominated by Scots pine and Norway spruce. The conditions for measurement were generally favorable due to good visibility and flat terrain.

Regular as well as control surveys were carried out by surveyors working alone. Regular surveyors had at least a 2-year forestry education. They knew that a part of their work would be subject to re-measurement on a random basis. Control surveyors were well educated and experienced, and were allowed to work under less time pressure than the regular surveyors. Having more time available, the control surveyors calibrated their electronic instruments more frequently than the regular surveyors. A control surveyor would also bring difficult increment cores back to the office for counting tree rings by aid of simple optical devices.

Regular surveyors and control surveyors used the same types of instruments for the measurements. The Forestor DME (digital distance measurer) and steel tape (for calibration) were used to establish the plot boundaries. Tree diameters were measured by a Forestor 500 data caliper. Tree heights were generally measured by simple non-digital instruments from Suunto and Silva. The Forestor Vertex electronic hypsometer was used by regular surveyors in 1997 and by control surveyors during 1994-1997. The reader is directed to Jonsson (1991) for details about the electronic instruments being used. Traditional increment borers were used for extracting cores for counting tree rings. Simple optical devices were available to aid in the tree-ring count. All data were recorded directly into hand-held computers.

The data set was adjusted for growth that might have occurred between the time for the regular survey and the time for control survey. The control measurements were always made during the same field season as the regular measurements.

Consultant, OL Forest Inventory LTD, Köpmangatan 44 A, 83133 Östersund, Sweden. E-mail: ola.lindgren@olskog.se

ANALYSIS

Strictly speaking, we are not able to analyze true measurement errors, i.e., the differences between the measurements being recorded and the true values. For example, it is almost impossible to determine the true value of the basal area on a sample plot. It would be extremely time-consuming to get indisputable ground level references for each tree. Consequently, we are instead looking at differences between measurements made by the regular surveyors as compared to control measurements. The recorded measurements are furthermore affected by both "genuine" measurement errors and recording errors.

We have reason to believe that the measurements made by the control surveyors are practically bias-free. However, even the control measurements will have a random error component of the same magnitude, or somewhat smaller, than the regular measurements.

The analysis and presentation of the data for each variable focuses on the differences in the paired observations

$$d_{ij} = rm_{ij} - cm_{ij}$$

where rm = regular measurement
cm = control measurement
i = regular surveyor being controlled
j = item (tree, plot) being controlled

Mean values and standard deviations of the differences were calculated for the whole data set (pooled data) and for individual surveyors. By storing the statistics of the differences for individual surveyors, new data sets were created. In such a database, each case (record) refers to one of the 87 regular surveyors. The analysis of surveyor performance adds significantly to the results of this study. One can assess the proportion of low-performing surveyors needing extra attention. The results also constitute a benchmark test by which one can evaluate other surveyors working under similar conditions.

The mean difference in surveyor level is a measure of surveyor "bias" in relation to the control, whereas the standard deviation of the differences is a measure of the random component of the measurement error. The given standard deviations should be divided by the square root of 2 to express an approximate standard deviation of measurement errors in relation to true values (assuming random errors of the same magnitude in the control data set, and assuming independence between the respective random error components).

It is obvious that surveyor bias is an important criterion of measurement quality. But what about the random error component, doesn't it cancel out anyhow if only a sufficient amount of data is collected? Yes, but a large random error component will tend to hide bias because the estimated mean difference will be of poor precision, and statistical tests against zero will not be powerful. Furthermore, large random errors in variables such as tree age might result in biased growth predictions when using a growth function containing tree age.

Before comparing the statistics between surveyors, the mean differences and standard deviations were "normalized." This was done by re-expressing the statistics (of the differences) in percent of the surveyor level averages (according to control) of the respective tree or plot variable. This was necessary since different surveyors worked under somewhat different conditions of average tree height, basal area, and other variables.

RESULTS

Tree Height

Pooled Data

The mean difference of the whole data set was –0.26 dm, or –0.16 percent of the average tree height (157.7 dm). The standard deviation of the differences was 8.47 dm (5.4 percent of average tree height). There was no statistically significant bias in the pooled height measurements. Frequency tabulation showed that 84.5 percent of the trees differed by less than 1 m. The distribution of differences had somewhat larger tails than the normal distribution.

Results varied among species. For Scots pine, the regular surveyors underestimated the height by 0.75 dm, a mean difference significantly apart from zero (99.9 percent). For other species, heights were, on average, slightly overestimated. The standard deviation was largest for white birch (9.8 percent) compared with 4.6 percent for Scots pine and 5.7 percent for Norway spruce.

Figure 1 shows the differences in tree height plotted over tree height. The relation is unexpected; large differences occur quite frequently, also for low trees. This might be due to observation errors (wrong scale) or to recording errors.

Surveyor Performance

Figure 2 shows the results by surveyors, each point showing the measurement statistics for one single surveyor. On the x-axis, mean differences are shown expressed in percent of the average measurement of each surveyor. The y-axis shows the standard deviations of the differences, also expressed in percent of the average measurement of each surveyor. Markers show if the mean differences passed a 99 percent significance test (t-test) against zero (square) or not (cross). The number of

Figure 1.—*Differences in total tree height plotted against the tree height according to the control.*

Figure 2.—*Evaluation graph for total tree height. Each marker shows the results for a single surveyor. Mean differences (x-axis) and standard deviations of the differences (y-axis) are expressed in percent of the surveyor level averages of tree height according to the control.*

observations behind each marker also influences the presence of statistical significance. The surveyors can roughly be divided into the following groups:

1. "The elite" is located close to the 0-line on the x-axis and in the lower part of the figure, indicating low bias and small random errors.

2. "Biased but consistent" are located away from the 0-line on the x-axis, but in the lower part of the figure. This group shows biased measurements, but because they are consistent it is possible to adjust for the bias, provided that a control survey is carried out.

3. "The tricky ones" are located in the upper part of the figure. A large random error makes it difficult to judge whether bias is present or not. If large bias is indicated, additional control data are often needed to determine if the members of this group need correction.

For tree height, 34 out of 87 surveyors showed significant (99 percent) differences compared to the control. The outcome was balanced: 18 had a negative mean difference and 16 had a positive mean difference. The magnitude of the bias is not large—on average, 4 dm or less than 3 percent of the average tree height. The massive appearance of statistically significant mean differences is due to the low random errors in height measurements. The surveyor with the largest bias had a mean difference of 7 percent (located in the upper right of the figure).

Tree Age at Dbh

Pooled Data

The mean difference was −1.47 years, or −1.8 percent of the average age of 80.8 years at breast height. The standard deviation was quite large, 14.5 years (17.9 percent of average age). The underestimation of tree age is statistically significant (>99.9 percent). A frequency study showed that only 75.5 percent of the trees differed by less than 10 years. The tails of the distributions were large compared to a normal distribution.

On the species level, statistically significant underestimation was found for Scots pine (-1.3 years) as well as for Norway spruce (-2.0 years). The standard deviation was very large for white birch (35.0 percent) compared to Scots pine (14.9 percent) and Norway spruce (19.6 percent). Those results are consistent with the difficulties experienced by the surveyors when counting tree rings from different species in the field.

The relation between the age differences and age was quite expected. Up to age 80, differences were reasonably small, at least for the conifers. Large differences for conifers are almost always tied to old trees (the oldest trees in the data set were about 300 years old).

Surveyor Performance

Figure 3 shows the troublesome results of the tree age comparisons at surveyor level. Here, the group "the tricky

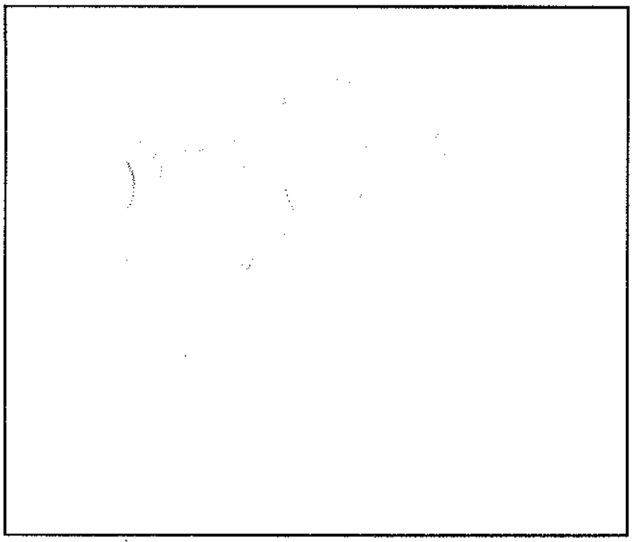

Figure 3.—*Evaluation graph for tree age (at dbh). Each marker shows the results for a single surveyor. Mean differences (x-axis) and standard deviations of the differences (y-axis) are expressed in percent of the surveyor level averages of tree age according to the control.*

ones" (cf. tree height above) is well represented. In the upper left and upper right of the figure, markers indicate surveyors with large random errors and possible large bias, but the bias indications are not statistically significant due to the large random errors. We do not have enough information to make the corrections that appear to be badly needed. In spite of the skewed distribution of mean differences, only 19 out of 87 mean differences are separated from zero with statistical significance. The magnitude of the bias (for the significant cases) is, on average, 6 years or about 9 percent.

Site Index

Pooled Data

Site indices for Scots pine and Norway spruce were estimated for each sample plot. The index H100 refers to the expected dominant height (in m) at total age 100. The most frequent method was to use site factors such as vegetation type, soil moisture, and elevation to estimate the index (Hägglund and Lundmark 1977). On plots with suitable conditions, site indices were also estimated with height development curves. Thus, the estimation of site indices required more of the subjective classification than other measurements described in this study. Consequently, the control measurements are also less "true."

The overall mean difference was 0.19 m, or 0.93 percent of the average index of 20.4 m. The standard deviation was 1.7 m (8.3 percent of average index). The slight overestimation was statistically significant. Frequency tabulation showed that over 80 percent of the plots differed by less than 2.0 m. A scatterplot showing the differences against the index itself (according to control) indicated that regular surveyors tended to overestimate low indices and to underestimate high indices.

Surveyor Performance

The evaluation plot in figure 4 shows a tight cluster of surveyors with non-significant mean differences and reasonable standard deviations. Some cases of significant bias are shown (15 out of 87 cases). Of those 15 cases, 13 have a positive mean difference, indicating overestimation of site index. The magnitude of the bias is just below 1 m, or 5 percent. A special analysis of the surveyors (with bias indication) showed that they sometimes had neglected index-reducing site conditions such as dry soil moisture or shallow soil depth.

Tree Count

Pooled Data

For this analysis, as well as for the following analysis of basal area, a filter was applied on the data set. Trees were defined as stems that are alive and having dbh of 5 cm or more. (In the actual data collection, both dead trees and trees below 5 cm were recorded.) Thus, differences in tree count are not only a result of classifying the trees as "in or out."

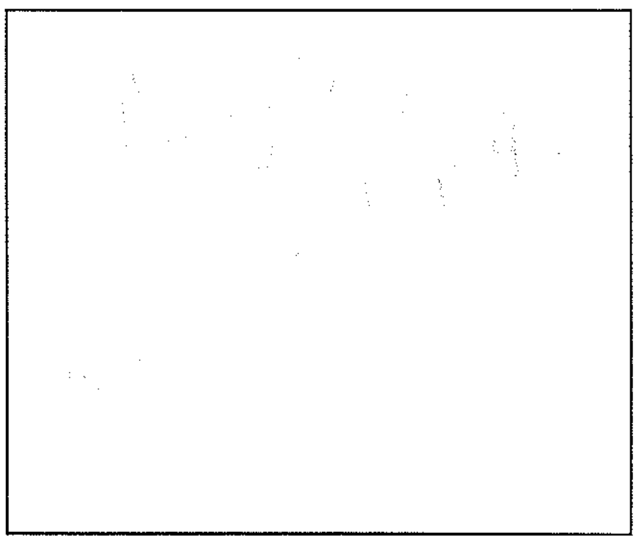

Figure 4.—*Evaluation graph for site index. Each marker shows the results for a single surveyor. Mean differences (x-axis) and standard deviations of the differences (y-axis) are expressed in percent of the surveyor level averages of site index according to the control.*

The mean difference was –0.52 trees per sample plot, or -2.8 percent of the average tally of 18.8 trees per plot. The standard deviation was 2.17 trees (11.5 percent of the average tally). The mean difference was separated from zero with very high statistical significance. If a difference in tree count is to occur, the sign of the difference is likely to be negative. Since each calipered tree is marked, cases of double-counting are rare. It is more likely to "forget" trees. According to the control surveyors, two types of trees are easy to forget: (1) trees near the plot boundary that surveyors are likely to classify as "out" without bothering to measure the distance, (2) trees very close to the plot center that are "left to take later," but are actually forgotten.

It is obvious that large differences in tree count require a certain number of trees on the plot. However, as soon as the tree number reaches 10, large differences begin to occur. Mean differences were computed separately for the different plot sizes applied (radius 5-10 m). The relation was not very strong. On the smallest (5 m) plots, 2.2 percent of the trees were missing (average count 12.4), whereas 3.0 percent of the trees were missing on the 10-m plots (average count 23.7).

Surveyor Performance

Figure 5 shows that the overall tendency to forget trees is not common for all regular surveyors. Most have a small mean difference, but the group with a statistically significant (99 percent) bias is quite large, 24 of 87 surveyors.

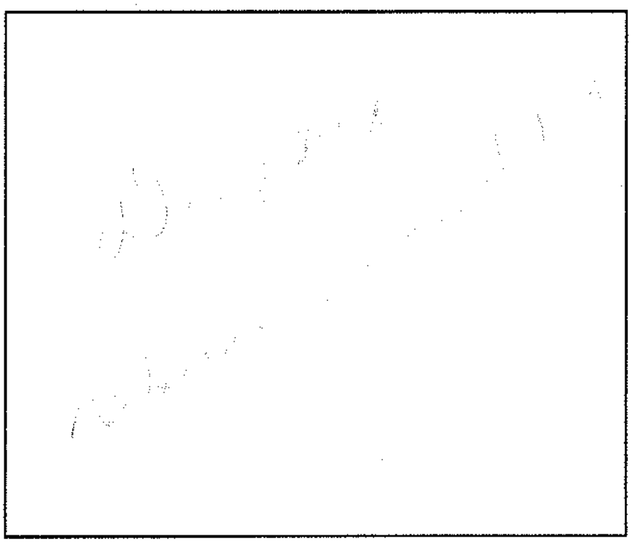

Figure 5.—*Evaluation graph for tree count. Each marker shows the results for a single surveyor. Mean differences (x-axis) and standard deviations of the differences (y-axis) are expressed in percent of the surveyor level averages of tree count according to the control.*

Of those 24 with bias, 21 have underestimated the number of trees. The average bias in this group of 21 underestimating surveyors is –1.6 trees (-8.0 percent), a rather bad result. However, cases of large underestimation are generally statistically significant. There are a few cases (squares in the upper left of the figure) with a strong tendency to forget trees, but where the statistical tests are not powerful enough to verify the possible bias.

Basal Area Per Plot

Pooled Data

The basal area is expressed as dm^2 per plot. The mean difference was –0.276, or -0.7 percent of the average basal area of 41.5 dm^2. The standard deviation was 3.56 dm^2 (8.6 percent of average basal area). The slight underestimation of basal area is statistically significant (99.9 percent).

Looking back to the results for the number of trees per plot, 2.8 percent of trees were missing, whereas only 0.7 percent of the basal area is missing. One explanation for this is that small trees are forgotten more frequently than large trees are. Another explanation for the different results in tree count and basal area is that regular surveyors are overestimating the diameters. Unfortunately, the data set doesn't contain any pairwise data sets of single tree diameters.

An effort was made to estimate differences in diameter measurement by trimming the plotwise data set. A subset was made by selecting plots with (1) the same number of trees according to regular and control surveyors and (2) a difference (absolute value) between basal areas that was smaller than 5 percent of the basal area itself. The second condition was introduced to try to eliminate plots with different trees measured, although the tree count was the same. After trimming, the data set consisted of 962 plots (out of 2,523). The mean difference in basal area in the trimmed data set was –0.11 dm^2, or just –0.3 percent of the average basal area. This difference was only close to statistical significance at the 95 percent level. Since there was no strong evidence of differences in diameter measurements, one might find it likely that the difference between the results for tree count and basal area has another explanation. The conclusion is then that small trees are more easily forgotten than large trees. The control surveyors, who often made notes about careless treatment of small trees, support this conclusion.

Surveyor Performance

The individual results of basal area measurements are shown in figure 6. Most of the surveyors performed well. Nine surveyors had statistically significant mean differences (in relation to zero), seven of whom underestimated

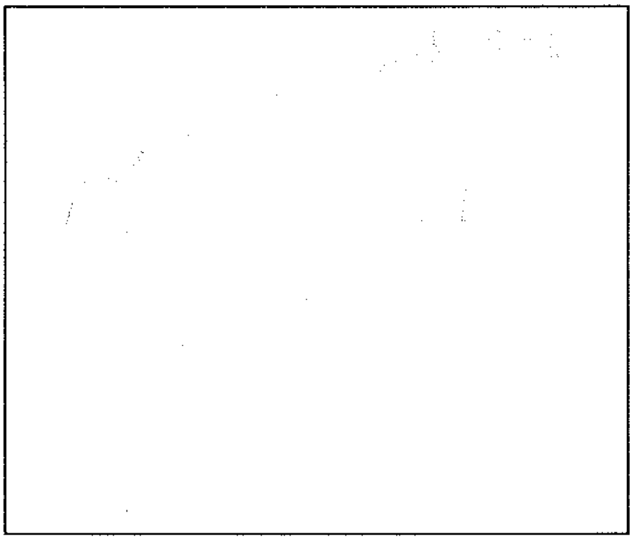

Figure 6.—*Evaluation graph for basal area per plot. Each marker shows the results for a single surveyor. Mean differences (x-axis) and standard deviations of the differences (y-axis) are expressed in percent of the surveyor level averages of basal area according to the control.*

the basal area. The average bias in this group of 7 is –5.8 percent, which is a serious level of bias.

The figure also shows quite a few surveyors with indication of bias, but this bias is not significant when exposed to a standard statistical test. More data would be needed to better evaluate this group.

Spatial Accuracy

The regular surveyors used a simple hand compass and a Haglöf Walktax distance measurer (thread) to lay out the sample plots. The distances from sample plots to roads or other distinct terrain features were within the range 0-300 m. However, the plots were not laid out independently. Plot 2 was usually laid out by measuring from plot 1 and so on.

A small study of spatial accuracy was made on 46 randomly selected control plots. The control surveyor was equipped with a Silva GPS recorder, using radio-transmitted real-time differential corrections giving a typical error of location of 2 m.

The average location error was 19 m, whereas the largest error was 41 m. The acceptable location error was specified to 30 m in the field manual. The study showed that 85 percent of the plots were located according to this norm.

The location error was positively correlated ($r=0.40$, $p=0.0056$) with the sequence number of sample plots within stands. Since the surveyors numbered their plots in consecutive order, this shows that the location errors are aggregating during the work within the stand. A significant positive correlation was also found between location error and the distance to road. However, some plots at roadside also had large location errors. This was found to be a result of misplacement of new roads on the map due to bad surveying practices.

CONCLUSIONS

The quality of field measurements made by different surveyors varied substantially. Serious levels of bias were present for about 10-20 percent of the group of 87 surveyors. In some cases, the bias was associated with small random errors, making it possible to calibrate the regular measurements. In other cases, the analysis indicated severe bias, but random errors were too big to make it possible to estimate the level of the bias with reasonable accuracy. The following conclusions can be drawn from this synthesis of surveyor performance:

1. Control surveys are justified. In some field projects, the bias resulting from measurements had a magnitude that was greater than the standard error due to sampling. In other field projects, the control survey only verified the good quality of the regular measurements. But that is also a valuable finding. The important thing is that the measurement errors are under control. Furthermore, the very existence of a control survey is judged to have a positive effect on the quality of field work.

2. There is a need for improved training of the surveyors before the work is started. Even experienced foresters have to be trained in the art of accurate measurement.

3. It should be possible to detect potential problems at an early stage of work. This is to avoid the situation when the "tricky cases" (indication of severe bias in connection with large random errors) are discovered after the end of the field season. Early detection will make it possible to give the surveyor additional training or at least to collect more control data before winter comes.

4. The "evaluation graphs" (figs. 2-6) can be used as a benchmark test when a first set of control data is analyzed. Depending on the outcome, the control fraction can be increased for the surveyors showing signs of inferior work.

Finally, some comments need to be made on the use of new instruments. The introduction of electronic hypsometers has not yet made any measurable effect on the quality of height measurements in our control program. Improving the ability to make accurate readings from a given location on ground is, of course, an advantage. However, the actual choice of location for making the measurement affects the final measurement quite a bit, as compared to the formal accuracy of reading as given in the technical specifications for the instruments. When electronic hypsometers are used, instrument-specific bias may occur due to erroneous calibration or other technical malfunctions. It might be easier to detect a malfunction in a simple device. It is strongly recommended that all kinds of hypsometers be checked on a regular basis by measuring objects with known heights, such as flagpoles or buildings.

Some preliminary trials comparing different types of calipers have shown substantial differences in diameter readings, even when the breast height location and the measurement direction are fixed. A routine analysis of a control survey made us aware of this problem.

ACKNOWLEDGMENTS

The author thanks the clients who made their field data available to this study and the personnel who did the field work. The following people reviewed this manuscript: Torleif Carlsson, The Forestry Research Institute of Sweden, Uppsala, Sweden; Göran Kempe, Swedish University of Agricultural Sciences, Department of Forest Resource Management and Geomatics, Umeå, Sweden; and Göran Ståhl, Swedish University of Agricultural Sciences, Department of Forest Resource Management and Geomatics, Umeå, Sweden.

LITERATURE CITED

Hägglund, B.; Lundmark, J.-E. 1977. Site index estimation by means of site properties. Studia Forestalia Suecica 138.

Jonsson, B. 1991. Electronic instruments and computerized devices in forest management planning. Report 23. Swedish University of Agricultural Sciences, Section of Forest Mensuration and Management. 20 p.

Jonsson, B.; Jacobsson, J.; Kallur, H. 1993. The forest management planning package. Theory and application. Studia Forestalia Suecica 189. 56 p.