Abstract
Tree-wise re-measured permanent sample plot data are needed to create and maintain modern forest growth models for Estonia. For that purpose an all-Estonian network of forest sample plots was established by the Department of Forest Management at Estonian Agricultural University in 1995. Tree species, storey, azimuth, distance from plot centre, two breast height diameters and faults were determined for each tree, and total height, crown base and height of dry branches were measured for sample trees. Routine inspection of forest plot measurement data for outliers was necessary because it provided information about the causes and consequences of measurement errors. At the moment, 715 forest sample plots have been established. A total of 749 outliers have been checked for errors on 112 plots in 2004.

In this study several statistical methods were tested for the detection of outliers. Empirical distributions of most tree variables by species and storey were analyzed using Grubb’s test, Dixon’s test and the 2-sigma rule. Multivariate methods (residual diagnostics) were used for detecting outliers in the interaction of several variables.

Analyzing empirical diameter distributions, Dixon’s test for small samples and 2-sigma rule for large samples were more effective than Grubbs’ test. The results of regression diagnostics of influential observations showed that Covratio method was the most sensitive one for outlier detection in nonlinear model and Dffits method in linear models. The most effective method for checking outliers from height-diameter data was the studentized residual.

Outlier detection plays important role in modeling, inference and even data processing because outliers can lead to model misspecification, biased parameter estimation and poor forecast.

Introduction
In forestry stand and tree-wise growth and yield modeling is a relevant question today. The key to the successful sustainable timber management is a proper understanding of the
growth processes, and one of the objectives of forest development modeling is to provide the tools that enable foresters to compare alternative silvicultural approaches. The current Estonian forest structure models are poor (Nilson, 1999). Both adequate forest stand descriptions and stand growth and structure models are needed for the effective use of the system.

The development of forest growth and yield based on stand-level models, diameter distribution models and individual tree models is a response to changing management objectives (Pretzsch et al., 2002). Growth modeling is also an essential prerequisite for evaluating the consequences of a particular management action for the future development of an important natural resource, such as a woodland ecosystem. Forest models represent average experience on how trees grow and how forest structures are modified (Gadow and Hui, 2001).

The accuracy of the growth and mortality models, is known, the interest usually lies in the development of stand volume (or other stand characteristics), the accuracy of which cannot be directly stated. It is very important to know the accuracy of predicted characteristics (Kangas, 1997). So the outlier detection plays an important role in modeling, inference and even data processing because outliers can lead to model misspecification, biased parameter estimation and poor forecasts (Tsay et al., 2000; Fuller, 1987)

Tree-wise re-measured permanent sample plot data are needed to create and maintain modern forest growth models for Estonia. As 715 forest sample plots have been established at the moment, routine inspection of forest plot measurement data for outliers is necessary because it provides information about the causes and consequences of measurement errors (Hordo, 2004). Statisticians recommend that data be routinely inspected for outliers, because outliers can provide useful information about the data. At the same time, methods for identifying outliers need to be used carefully, as it is quite easy to confuse discordant random observations and outliers (Iglewicz and Hoaglin, 1993; Carroll et al., 1995). In the study the term ‘outlier’ means an observation that has a substantial difference between its actual and predicted dependent variable (a large residual) values or between its independent variable values and those of other observations. The objective of denoting outliers is to identify observations that are inappropriate representations of the population from which the sample is drawn, so that they may be discounted or even eliminated from the analysis as unrepresentative (Hair et al, 1998).

The aims of this paper are:

- Overview of several tests for the detection of outliers from permanent sample plot data;
Comparison and estimation of efficiency of different outlier tests for detecting measurement errors.

Material and methods

**Network of permanent sample plots**

Data on the network of permanent forest growth plots were used. A method of establishing a network of sample plots was developed at the Department of Forest Management of the Estonian Agricultural University. The method of establishing permanent forest growth plots is mainly based on the experience of the Finnish Forest Research Institute (Gustavsen et al., 1988). Some recommendations from Curtis (1983) were taken into consideration.

In this study data on 715 sample plots were ascertained. On each plot, measurements of more than 100 trees were recorded in 1995-2004. The check program was applied on 715 sample plots and was re-measured on 112 sample plots in the summer 2004. Each forest element (storey and tree species) on sample plot was checked as independent sample.

Tree species, storey, azimuth, distance from the plot centre, two perpendicular breast height diameters and faults were measured for each tree, and total height, crown length and height of dry branches were measured for model trees (Kiviste, Hordo, 2003).

**Analysis**

In the study, several methods were used to detect outliers for the purpose of finding out measurement errors in the permanent sample plot database arising from data collecting, entering or from wrong coding. In this study, authors were focused on outlier detection methods, which may be distinguished as follows: distribution base tests and regression diagnostics.

**Distribution based tests**

Distribution based tests were used for detecting outlier from diameter distribution and were applied all trees. Empirical distributions of most tree variables by species were summarized and several statistical criteria were elaborated to detect outliers in the data set. All the methods first quantify how far the outlier is from the other values. This can be the difference between the outlier and the mean of all points, the difference between the outlier and the mean of the remaining values, or the difference between the outlier and the next closest value (Grubbs, 2003).

The following tests were used to detect outliers/measurement error from diameter distribution:
- Grubbs’ test is used to detect outliers in a univariate data set (sample $N > 3$). It is based on the assumption of normality. That is, you should first verify whether your data could be reasonably approximated by a normal distribution before applying the Grubbs’ test. The Grubbs’ test statistic is defined as: 
  \[ G = \frac{\max(\bar{d} - d_i)}{s_d} \]  
where $\bar{d}$ and $s_d$ are the sample mean and standard deviation. The Grubbs test statistic is the largest absolute deviation from the sample mean in units of the sample standard deviation (Motulsky, 1999; Grubbs, 2003). Test statistic was decided in terms of the bound given in the table by Grubbs (2003).

- Dixon’s test (sample $N \geq 3...25$) is used to find out whether the mean ($\mu$) of some data set differs substantially from the $k - 1$ mutually different mean of other data sets, ordering them by magnitude - in increasing order if the mean in question is the smallest, in descending order if it is the largest. The significance bounds depended on the sample $N$. Then the test statistic 
  \[ D = \left| \frac{d_1 - d_2}{d_1 - d_{\max}} \right| \]  
was computed (e.g., for $3 \leq N \leq 7$) and decided in terms of the bound given in the table by Sachs (1982).

- The “2-sigma region” was recommended by Sachs (1982), Sheskin (2000), Iglewicz and Hoaglin (1993) for sample size larger than $N = 25$ the extreme values and was tested by means of the test statistic, 
  \[ T = \left| \frac{x_i - \mu}{\delta} \right| \]  
where $x_i$ is the supposed outlier and $\mu$ is mean-value and $\sigma$ is standard deviation. The value is discarded as an outlier if $T$ equals or exceeds the region $\mu \pm 2 \sigma$, where the mean and standard deviation are computed without the value suspected of being an outlier.

Residual diagnostics – height-diameter check
An assumption of regression is that the predictor variables are measured without errors. Several residual diagnostics were for detecting outliers from height – diameter dataset used and were applied on sample trees dataset. Two regression (height curve) models were used to detect height measurement errors, outliers from sample trees:

- logarithmic function  
  \[ H = a + b \cdot \ln(D) \];
- simple linear function  
  \[ H = a + b \cdot D \].

Regression diagnostics were developed to measure various ways in which a regression relation might derive largely from one or two observations. Observations whose inclusion or exclusion result’s in substantial changes in the fitted model (coefficients, fitted values) are said to be influential (Hair et al., 1998; Tsay et al., 2000; Hordo, 2004). Outliers in the
response variable represent model failure. The following methods are used for detecting outliers from sample plots:

- **Leverage method** 
  \[ h_i = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{v^2} \]
  where \( v^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 \), \( n \) - sample size, \( x_i \) - \( i \)-th observation, \( \bar{x} \) - mean. Outlier was detected if \( h_i > 0.5 \). Leverage - an observation that has substantial impact on the regression results in its differences from other observations on one or more of the independent variables. The most common measure of the leverage point is the hat value, contained in the hat matrix (Hair et al., 1998; Iglewicz and Hoaglin, 1993).

- **Standardized residual** 
  \[ s_i = \frac{e_i}{\hat{s} \sqrt{1 - h_i}} \]
  where \( e_i \) – residual, \( \hat{s} \) - standard deviation of residuals, \( h_i \) – leverage of \( i \)-th observation. Outlier was detected then \( s_i > \pm 2 \). Standardized residuals are the residuals divided by the estimates of their standard errors. Their mean is 0 and standard deviation is 1. There are two common ways to calculate the standardized residual for the \( i \)-th observation. One uses the residual mean square error from the model fitted to the full model (internally studentized residuals). The other uses the residual mean square error from the model fitted to all the data except the \( i \)-th observation (externally studentized residuals). The externally standardized residuals follow a t-distribution with \( n-p-2 \) df. They can be thought of as testing the hypothesis that the corresponding observation does not follow the regression model that describes the other observations (Hair et al., 1998; Iglewicz and Hoaglin, 1993).

- **Studentized residual** 
  \[ Stud = r_i = \frac{e_i}{\hat{s}(i) \sqrt{1 - h_i}} \]
  where \( e_i \) – residual, \( \hat{s}(i) \) - standard deviation of residuals, \( h_i \) – leverage of \( i \)-th observation. Outlier was detected then \( Stud > \pm 2 \) (Hair et al., 1998; Iglewicz and Hoaglin, 1993).

- **Cook’ Distance** 
  \[ D_i = \frac{r_i^2 \cdot h_i}{p \cdot (1 - h_i)} \]
  where \( r_i \) – studenized residual, \( p \) – number of parameters, \( h_i \) – leverage of \( i \)-th observation. Outlier was detected then \( D_i > 2/\sqrt{n} \). Cook’s Distance for the \( i \)-th observation is based on the differences between the predicted responses from the model constructed from all the data and the predicted responses from the model constructed by setting the \( i \)-th observation aside. For each observation, the sum of squared residuals is divided by \((p+1)\) times the Residual Mean Square from the full model (Becker, 2000; Hair et al., 1998).
• Dffits test \( F_i = \frac{\hat{\mu}_i - \hat{\mu}_{(i)}}{s_{(i)} h_i} \), where \( \hat{\mu}_i \) - predicted value for \( i \)th observation, \( \hat{\mu}_{(i)} \) - predicted value for \( i \)th observation without \( i \)th observation, \( s_{(i)} \) - standard deviation of residual without \( i \)th observation. Outlier was detected then \( |F_i| > \frac{2}{\sqrt{n}} \). Dffits – measure of an observation’s impact on the overall model fit. Dffits is the scaled difference between the predicted responses from the model constructed from all the data and the predicted responses from the model constructed by setting the \( i \)th observation aside (Hair et al., 1998).

• Covratio test \( C_i = \left| \frac{s_{(i)}^2 (X_{(i)}X_{(i)})^{-1}}{s^2 (X'X)^{-1}} \right| \), where \( X_{(i)} \) - observation matrix without the \( i \)th observation, \( s_{(i)} \) - standard deviation of residuals without the \( i \)th observation. Outlier was detected then \( |C_i - 1| > \frac{3p}{n} \). Covratio – measure of the influence of a single observation on the entire set of estimated regression coefficients (Hair et al., 1998).

• Dfbetas test \( DFBETAS_{j,i} = \frac{\hat{\beta}_j - \hat{\beta}_{(j)}(i)}{\sqrt{s_{(i)}^2 C_{jj}}} \), where \( \hat{\beta}_j \) - regression coefficient without \( i \)th observation, \( \hat{\beta}_{(j)}(i) \) - regression coefficient with \( i \)th observation, \( s_{(i)} \) - standard deviation of residuals without the \( i \)th observation, \( C_{jj} \) - diagonal element of Covratio matrix. Outlier was detected then \( |DFBETAS_{j,i}| > \frac{2}{\sqrt{n}} \). Dfbetas – measure of the change in a regression coefficient when an observation is omitted from the regression analysis. Dfbetas are similar to Dffits. Instead of looking at the difference in fitted value when the \( i \)th observation is included or excluded, Dfbetas looks at the change in each regression coefficient (Hair et al., 1998).

Results
A Visual FoxPro program was elaborated to detect measurement errors and outliers in the forest sample plot data. In 2004 total of 112 sample plots were re-measured. Using data measured in 1999 on those 112 sample plots, totally of 749 outliers were detected by the test program. Applying at the same time by test program several (distribution and residual diagnostic based) outliers detection methods were detected more than one outlier on tree. During re-measurement were outlier trees carefully checked, was it measurement error or not. Additionally were detected measurement errors by measurer, what the test program was
not able to detect. Totally 415 measurement errors of diameter (Table 2) were detected by re-measurement and 508 measurement errors of height (Table 3).

**Distribution based tests**

Grubbs test is universal for outlier detection for small (Figure 1) and large sample (Figure 2); however only one observation from sample size can have detected as outlier. In the study each forest element was checked as sample.

As an example, in the Figure 1 is shown visual picture of sample plot number 129, first storey pine. In the Figure 1 was sample of five diameters tested and the maximum diameter was as outlier detected, but actually it was correct observation. In the Figure 2, according Grubbs’ test for sample size \( n > 50 \), also maximum diameter was detected as outlier, but that was actually measurement error. By Grubbs test was detected 43 outliers (Table 2), and only three measurements were detected as measurement errors from outliers.

*Figure 1.* Sample plot number 129 and sample size \( n = 5 \) then \( G > 1.71 \); with dot ° is marked outlier (not measurement error).

*Figure 2.* Sample plot number 162 and sample size \( n > 50 \) then \( G > 3.61 \); with cross + is marked outlier/measurement error.
Dixon’s test can be used for smaller sample ($3 \leq n \leq 25$), and more than outlier can be detected. As an example, in the Figure 3, all Dixon’s test statistics values above the critical line ($\text{Dixon} > 0.5$) were outliers, and in this case it is also measurement error. By Dixon’s test were 33 outliers and only 5 diameter measurement errors of outliers detected (Table 2).

![Figure 3](sample_plot_175_aspen.png)

*Figure 3.* Sample plot number 175 and sample size $n = 15$ then $T > 2 \sigma$; with cross + is marked outlier/measurement error.

2-sigma rule can be used for large sample ($n \geq 25$). As an example, in the Figure 4, outliers were detected and they all were measurement errors. In addition, four other measurement errors were found during the re-measurement, but test was not able to detect them. By 2-sigma rule were 74 outliers and only 19 diameter measurement errors of outliers detected (Table 2).

![Figure 4](sample_plot_129_norway_spruce.png)
Figure 4. Sample plot number 129 and sample size $N > 25$ then $D > 0.525$; with crosses + are marked outlier/measurement errors, and with × are measurement errors (not detected as outliers).

Table 1 shows a comparison of results from the diameter distribution tests. Regarding the detection of measurement errors for diameter distribution, Dixon’s method (suitable for small samples) and the 2-sigma method (suitable for bigger samples) detected more outliers than Grubb’s test. Grubbs’ test detects one outlier at a time. This outlier is expunged from the dataset and the test is iterated until no outliers are detected.

Table 1. Number of diameter distribution outliers on permanent sample plots excluded from the different tests. 0 – no outliers detected, 1 – outliers detected

<table>
<thead>
<tr>
<th>Sample N 3...25</th>
<th>Grubbs’ test</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dixon’s test</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>3810</td>
<td>27</td>
</tr>
<tr>
<td>1</td>
<td>17</td>
<td>16</td>
</tr>
<tr>
<td>Total</td>
<td>3827</td>
<td>43</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample N &gt; 25</th>
<th>Grubbs’ test</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-sigma rule</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>15431</td>
<td>37</td>
</tr>
<tr>
<td>1</td>
<td>65</td>
<td>9</td>
</tr>
<tr>
<td>Total</td>
<td>15496</td>
<td>46</td>
</tr>
</tbody>
</table>

Results from diameter distribution are presented in Table 2. Grubb’s test is inefficient, method were 43 outliers detected, but only three were in the same time measurement errors. Dixon’s test for smaller and 2-sigma region for larger sample size seems to be better.

Table 2. Results from diameter distribution control.

<table>
<thead>
<tr>
<th>Method</th>
<th>Number of outliers</th>
<th>Number of measurement errors from outliers</th>
<th>Total number of measurement errors</th>
<th>Number of measurement errors from outliers</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dixon’s test (3 ≤ n ≤ 25)</td>
<td>33</td>
<td>5</td>
<td>107</td>
<td>5</td>
<td>4.7</td>
</tr>
<tr>
<td>2-sigma region (n ≥ 25)</td>
<td>74</td>
<td>19</td>
<td>308</td>
<td>19</td>
<td>6.2</td>
</tr>
<tr>
<td>Grubb’s test (n &gt; 3)</td>
<td>43</td>
<td>3</td>
<td>415</td>
<td>3</td>
<td>0.7</td>
</tr>
</tbody>
</table>

Height – diameter control
A total of 7580 height-diameter measurement from sample trees were analyzed. In the study on two regression models – linear and logarithmic – for outlier detection (using residual diagnostics) was studied. For each sample as forest element (storey and tree-wise) was predicted model parameters and was used in residual diagnostic. For residual diagnostic were used several methods. In this study, for checking conformity between outlier methods and measurement errors were used chi-square test.

As an example, in the Figure 5 shows visually, detected outliers with leverage method lies above the line. This presents all observations (sample trees) from the dataset. Leverage method shows influential observations. According Table 3 and Table 4 by leverage method were detected 149 outliers using linear model and 73 outliers using logarithmic model. Chi-square was higher for function of linear regression (Table 3) 39.6 and for function of logarithmic function only 0.0003.

![Figure 5. Leverage method, according linear (left) and logarithmic (right) model.](image1)

In the Figure 6, as an example, are presented standardized residuals, outside the lines (criteria ± 2) are outliers. For next test, in the Figure 7, studentized residuals also observations outside the lines are outliers, used for all observations. From linear model were detected 189 outliers with standardized residual test and from logarithmic model 325 outliers. According Table 3 and Table 4 the chi-square were almost equal, respectively 85.2 and 83.2.

![Figure 6. Standardized residuals, according linear (left) and logarithmic (right) model.](image2)
Figure 7. Studentized residuals –according linear (left) and logarithmic (right) model.

Cook’s Distance method was used also for linear and logarithmic models (Figure 8). In this case critical value depends on sample size. In the Figure 8 shows, using linear model for Cook’s D testing, two outliers were detected, they both occur measurement errors. Additionally were detected during re-measurement four measurement errors what were not outliers. And using logarithmic model was detected only one outlier/measurement error.

Figure 8. Cook’s Distance method, from linear (left figure) and logarithmic (right figure) models (with cross + is marked outlier/measurement error and with × measurement error, not detected as outlier).

Table 3. Results from height control, function of linear regression were used.

<table>
<thead>
<tr>
<th></th>
<th>Number of outliers</th>
<th>Number of measurement errors from outliers</th>
<th>%</th>
<th>Total number of measurement errors</th>
<th>Number of measurement errors from outliers</th>
<th>%</th>
<th>Chi-sq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leverage</td>
<td>149</td>
<td>29</td>
<td>19.5</td>
<td>508</td>
<td>29</td>
<td>5.7</td>
<td>39.6</td>
</tr>
<tr>
<td>Standardized residual</td>
<td>189</td>
<td>44</td>
<td>23.3</td>
<td>508</td>
<td>44</td>
<td>8.7</td>
<td>85.2</td>
</tr>
<tr>
<td>Studentized residual</td>
<td>258</td>
<td>67</td>
<td>26</td>
<td>508</td>
<td>67</td>
<td>13.2</td>
<td>158.6</td>
</tr>
<tr>
<td>Cook’s D</td>
<td>402</td>
<td>87</td>
<td>21.6</td>
<td>508</td>
<td>87</td>
<td>17.1</td>
<td>151.5</td>
</tr>
<tr>
<td>Covratio</td>
<td>676</td>
<td>80</td>
<td>11.8</td>
<td>508</td>
<td>80</td>
<td>15.7</td>
<td>31.3</td>
</tr>
<tr>
<td>Dffits</td>
<td>15</td>
<td>3</td>
<td>20</td>
<td>508</td>
<td>3</td>
<td>0.6</td>
<td>4.3</td>
</tr>
</tbody>
</table>
From analyzed several residual diagnostic methods, Table 3 and Table 4 show the results of residual diagnostic from height-diameter check. Comparing linear and logarithmic model with chi-square test, a result show that on both cases the studentized tests were best. The results of regression diagnostics of influential observations show that the Covratio method was the most sensitive one for outlier detection. Covratio method was detected the highest number of outliers 676 and 655, respectively from linear and logarithmic regression

Table 4. Result from height control, function of logarithmic regression was used.

<table>
<thead>
<tr>
<th>Method</th>
<th>Number of outliers</th>
<th>Number of measurement errors from outliers</th>
<th>%</th>
<th>Total number of measurement errors</th>
<th>Number of measurement errors from outliers</th>
<th>%</th>
<th>Chi-sq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leverage</td>
<td>73</td>
<td>5</td>
<td>6.8</td>
<td>508</td>
<td>5</td>
<td>1.0</td>
<td>0.003</td>
</tr>
<tr>
<td>Standardized residual</td>
<td>325</td>
<td>62</td>
<td>19.1</td>
<td>508</td>
<td>62</td>
<td>12.2</td>
<td>83.2</td>
</tr>
<tr>
<td>Studentized residual</td>
<td>447</td>
<td>79</td>
<td>17.7</td>
<td>508</td>
<td>79</td>
<td>15.6</td>
<td>91.4</td>
</tr>
<tr>
<td>Cook's D</td>
<td>124</td>
<td>17</td>
<td>13.7</td>
<td>508</td>
<td>17</td>
<td>3.3</td>
<td>9.9</td>
</tr>
<tr>
<td>Covratio</td>
<td>655</td>
<td>41</td>
<td>6.3</td>
<td>508</td>
<td>41</td>
<td>8.1</td>
<td>0.224</td>
</tr>
<tr>
<td>Dffits</td>
<td>11</td>
<td>3</td>
<td>27.3</td>
<td>508</td>
<td>3</td>
<td>0.6</td>
<td>7.5</td>
</tr>
<tr>
<td>Dfbetas</td>
<td>574</td>
<td>80</td>
<td>13.9</td>
<td>508</td>
<td>80</td>
<td>15.7</td>
<td>52.0</td>
</tr>
</tbody>
</table>

Discussion

Many statistical techniques are sensitive to the presence of outliers. Checking for outliers should be a routine part of any data analysis. Potential outliers should be examined to see if they are possibly erroneous. If a data point is in measurement error it should not be deleted without careful consideration.

In general the outlier identification procedure presumes data from normal distribution, as appears from a comparison of Grubbs' test with Dixon's test and 2-sigma region method. For Grubbs test it is highly recommended that result data come from normal distribution (Iglewicz, Hoaglin, 1993). According to Sachs (1982) Dixon’s test is rather prone to deviations from normality and variance homogeneity, since by the central limit theorem, the means of non-normally distributed data sets are themselves approximately normally distributed.

In practical situations diameter data are highly skewed, usually to the right. As an example, Figure 9 shows the tree mean diameter distribution is a unimodal on sample plots; distribution is skewed to right (skewness 0.996). According to Iglewicz and Hoaglin (1993) in such situations, a lognormal distribution is frequently more appropriate than a normal distribution.
Height-diameter regression models are more complicated, the relationships in models are not linear. The various diagnostic measures aim at enabling the user of least-squares regression to cope with the fact that all observations, no matter how discrepant, have an impact on the regression line (Iglewicz and Hoaglin, 1993). Outliers in (height-diameter) data may take more form than in univariate data, and detecting them offers more challenges. The situation is more challenging because some (height-diameter) data points have more impact than others on the fitted regression line. Iglewicz and Hoaglin (1993) have recommended specialized diagnostic techniques, and to use robust methods, instead of least squares.

Carroll et al. (1995) have summarized some of the known results about the effect of measurement error in linear regression, but a comprehensive account of linear measurement error models can be found in Fuller (1987). Usually focusing on simple linear regression and arriving at the conclusion that the effect of measurement error is to bias the slope estimate in the direction of 0 (Carroll et al., 1995).

Conclusions
At the moment, 715 forest sample plots have been established in Estonia. A total of 749 outliers have been checked for errors on 112 plots in 2004.

Measurement error enters into forestry in many different forms. The errors can have very negative effects on model parameters, model estimates, and the variances of model parameters and model estimates.

In this study several statistical methods were tested for outlier detection. Empirical distributions of most tree variables by species and storey were analyzed using Grubb's test, Dixon's test and the 2-sigma region method. Multivariate methods (residual diagnostics for

Figure 9. Diameter distribution of 91 495 trees on the 715 permanent sample plots.
linear and nonlinear models) were used for detecting outliers in the interaction of several variables.

Analyzing empirical diameter distributions, Dixon’s test for small samples and 2-sigma rule for large samples are more effective than Grubbs’ test.

The results of regression diagnostics of influential observations showed that Covratio method was the most sensitive one for outlier detection in nonlinear model and Dffits method in linear models. The most effective method for checking outliers from height-diameter data was studentized residual.

In this study, in the outlier analysis it was discovered that most of the outliers tested and put out by the control program were caused by stand natural disturbance regime, like mortality, competition, spatial location, diseases, effect of weather (frost damage, temperature), etc. As well, a large amount of the detected outliers in the changes of tree dimensions was caused by ungulate herbivores (like elk – eating spruce bark), effect of the harvest (intensity) as well as the effect of deforestation, understory, etc.

Outlier detection plays important role in modeling, inference and even data processing because outliers can lead to model misspecification, biased parameter estimation and poor forecast.

Acknowledgements
The establishment of the network of permanent forest growth plots was supported by the Estonian Science Foundation (Grants No 4813 and 5768), the Estonian State Forest Management Centre and the Centre of Environmental Investments.

References


