

UNCERTAINTY ANALYSIS OF STANDING VOLUME DATA STORED IN THE NATIONAL FOREST DATABASE

SUPPLEMENT TO THE NATIONAL GREENHOUSE GAS INVENTORY OF HUNGARY

*by: Tamás Tobisch and Péter Kottek
Department of Statistics and Data Analysis
Forestry Directorate
National Food Chain Safety Office*

Budapest, 2013

Table of contents

1.	Introduction	3
2.	Sampling errors in surveys.....	3
3.	Combined errors from sampling and applying yield tables	7
4.	Uncertainty of NFD area-specific volume values	10
5.	Uncertainty of forest subcompartment areas	12
6.	Uncertainty of standing volume by species groups on country level.....	13
7.	References.....	14

1. Introduction

In the Hungarian National Forest Database (NFD) standing volume data by tree species are stored on forest subcompartment level. These data are gained by multiplying an area- and species-specific volume (m^3/ha) data by the area of the given subcompartment. The former is most often taken from yield tables, although other surveying methods are also used (a detailed description of how NFD operates is available on the webpage: http://www.nebih.gov.hu/szakterulek/szakterulek/erdeszeti_igazgatosag/supplementary_inf_ERT/forest-db.html). Data from the yield tables are retrieved based on measurements (i.e. age, height, basal area) established in surveys during forest planning. Thus, the uncertainty of the m^3/ha values originates from two major sources: 1. sampling for the above measures during surveys; 2. modelling by yield tables.

2. Sampling errors in surveys

Two of the nine stand assessment methods that can be applied for sampling tree stands for standing volume and that are most frequently used in forest surveys are: basal area sampling and forecasting using yield tables. As the former is a systematic sampling, random errors of average values can be estimated by statistical methods. However, it should be noted that for height measurements, which is needed as inputs for the yield tables, trees are chosen preferentially. Consequently, the calculated uncertainty of average height does not reflect the true uncertainty of the volume data, which is presumably higher than the calculated one. In order to avoid underestimation of the real uncertainty we: (1) use a conservative approach and (2) compare the results with a former expert judgement (Czirok and Szabó 2011). At the same time, the uncertainty of average basal area values can be assessed correctly by standard statistical methods.

In order to quantify the confidence intervals of average values of sampling data we analysed data of the survey of 642 stands. This contained 26,353 measured basal area data, as well as measured height of 12,993 trees. From these data, percentage confidence intervals (with 95 % confidence levels) as well as relative standard deviations of distributions of mean values were calculated by the adjusted bootstrap percentile method which corrects of bias and skewness of the bootstrap distribution (the applied formula can be found in Chapter 5 of Davison, A.C. and Hinkley, D.V. 1997).

Confidence interval and standard deviation data over mean values are reported in Figures 1-3. In the figures each point represents a forest subcompartment. Data of sessile oak are shown only, but the results were similar for other species groups studied.

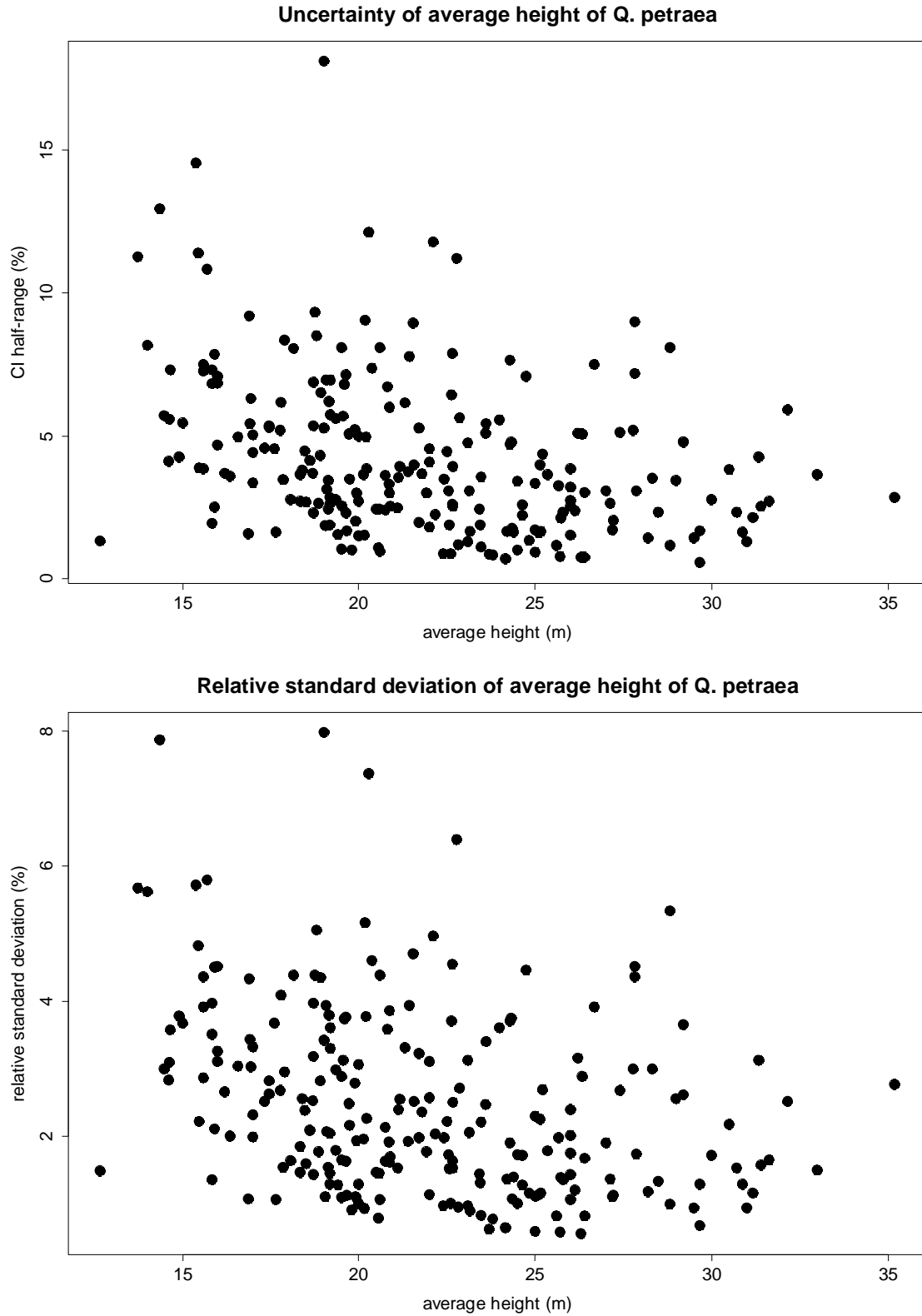


Figure 1. Results of bootstrap analysis I.

According to the results of bootstrap analysis, the uncertainty of average height is within $\pm 10\%$ in most cases (which means app. 5% relative standard deviation; Figure 1). This result is consistent with the above-mentioned expert judgement (Czirok and Szabó 2011), which assessed random errors of 0.5 m , 1 m and 2 m in stands shorter than 15 m , between 15 m and 30 m and taller than 30 m , respectively. In order to be conservative, the highest random error

of $\pm 10\%$ was assumed throughout our investigation, independently from the average height or tree species. The results of bootstrap analysis of the relative sampling errors by species are shown in Tables 1 and 2 for species with sample size of minimum 500 trees.

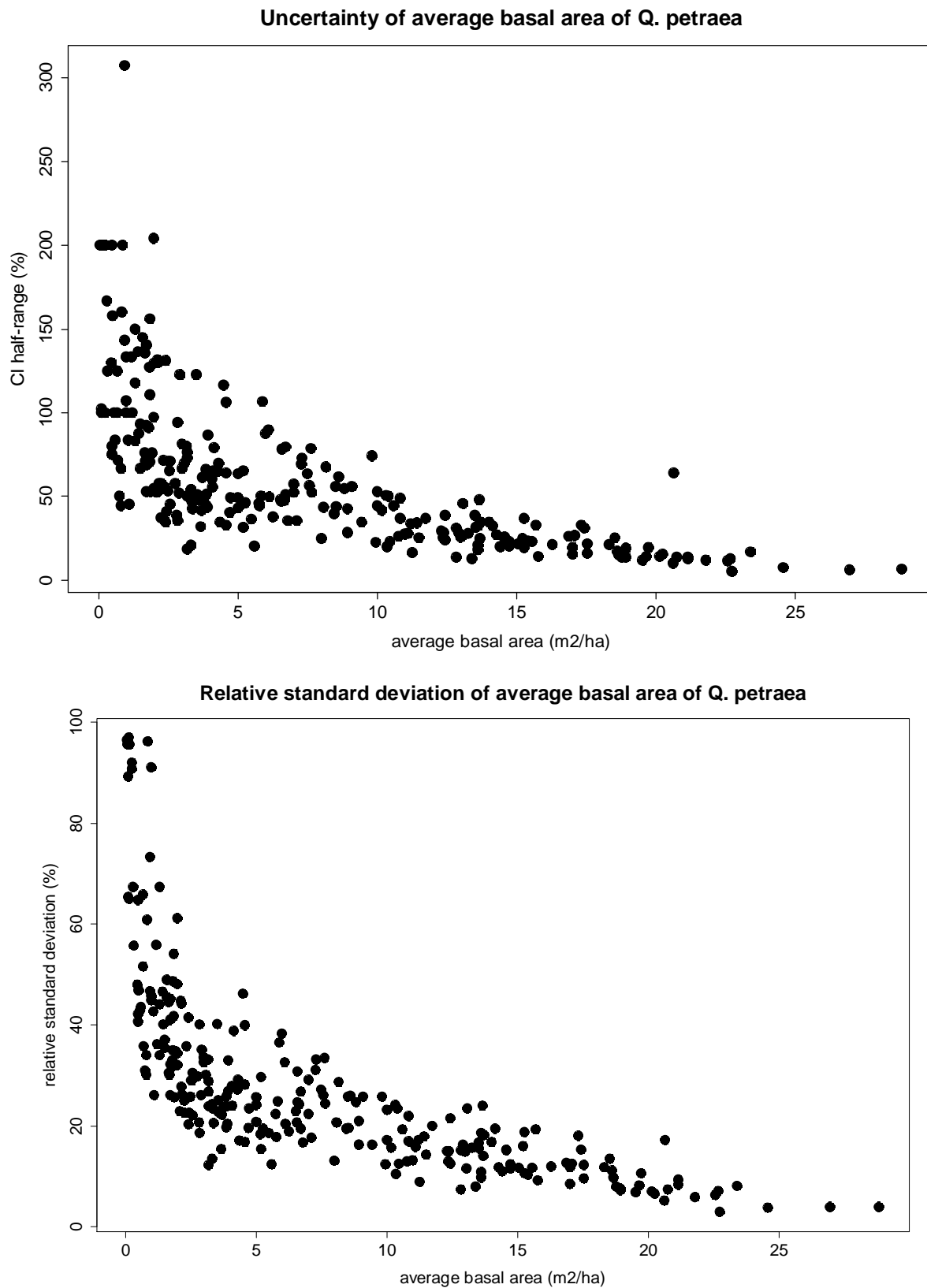


Figure 2. Results of bootstrap analysis II.

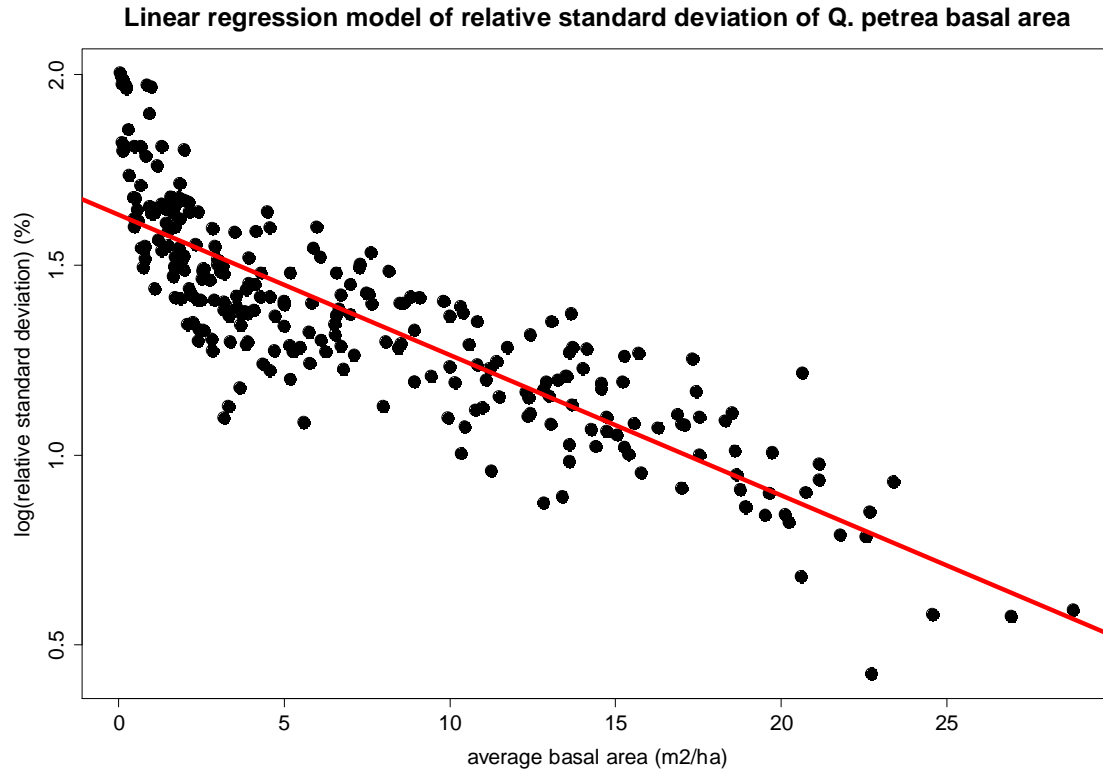


Figure 3. Linear regression between average basal area and logarithmically transformed relative standard deviation.

Table 1. Relative standard errors of average height of species groups with a sample size (no. of measured height data) larger than 500. The relative standard errors are the means of values calculated on forest subcompartment level.

species group	relative standard error (%)	sample size
<i>Quercus robur</i>	2.0	654
<i>Quercus petraea</i>	2.5	2386
<i>Quercus cerris</i>	2.5	2543
<i>Fagus silvatica</i>	2.2	2224
<i>Carpinus betulus</i>	3.6	1605
<i>Robinia pseudoacacia</i>	3.3	536

Table 2. Relative standard errors of average basal area of species groups with a sample size (no. of raw data) larger than 500. The relative standard errors are the means of values calculated on forest subcompartment level.

species group	relative standard error (%)	sample size
<i>Quercus robur</i>	24.3	2071
<i>Quercus petraea</i>	27.6	3694
<i>Quercus cerris</i>	27.1	4296
<i>Fagus silvatica</i>	28.3	3270
<i>Carpinus betulus</i>	38.8	3776
<i>Robinia pseudoacacia</i>	32.4	714
<i>Acer</i> spp.	53.3	509
<i>Fraxinus</i> spp.	45.3	738
Other hardwood species	71.2	531
<i>Tilia</i> spp.	50.4	984
<i>Pinus silvestris</i>	28.4	747

The uncertainty of average basal area strongly depends on its absolute value. In the case of species, the proportion of which is low relative standard deviation of basal area can be extremely high within the sample, because of patchy spatial distribution or because the species occurs in some sample plots and is absent from others. We created linear regression models between mean basal area and logarithmically transformed relative standard deviation for those species groups for which we received sampling data enough for quantifying uncertainties (Figure 3, Table 3), due to the strong relationship between the absolute value of the mean and uncertainty of basal area. For those species groups for which we did not receive enough data we applied an ‘average model’ which was based on the received data regardless of species.

Table 3. Parameters of linear regressions between mean basal area and logarithmically transformed relative standard deviation. The last row shows data of regression carried out on combined data of all species groups with sample size larger than 500.

species group	intercept	slope	R ²	p
Quercus robur	1.645525	-0.039169	0.71	< 2.2E-16
Quercus petraea	1.632143	-0.036811	0.73	< 2.2E-16
Quercus cerris	1.645746	-0.036704	0.76	< 2.2E-16
Fagus silvatica	1.693363	-0.035004	0.75	< 2.2E-16
Carpinus betulus	1.717897	-0.048225	0.62	< 2.2E-16
Robinia pseudoacacia	1.831256	-0.060009	0.86	< 2.2E-16
Acer spp.	1.820377	-0.082147	0.80	6.19E-14
Fraxinus spp.	1.757199	-0.050435	0.61	5.94E-10
Pinus silvestris	1.780258	-0.039922	0.91	< 2.2E-16
combined	1.698925	-0.0403565	0.75	< 2.0E-16

However, the linear regression models underestimate the true uncertainty for small values of average basal area (Figure 3). In order to avoid underestimation, we used the maximum of the calculated uncertainty values of the given species group for basal areas smaller than 1 m²/ha.

3. Combined errors from sampling and applying yield tables

The error of estimates of the yield tables have not been reported in literature, but can be assessed using measured volume data of the *Forest Growth Monitoring System* (GMS). This system was established to monitor changes in standing volume of forests and to verify the estimates obtained using yield tables. A detailed description of this project and its results is available in English on the webpage:

http://www.nebih.gov.hu/szakteruletek/szakteruletek/erdeszeti_igazgatosag/erdeszet_szakteruletek/monitoring/EMMRE_20_eve/fmo-eng.html.

The GMS is a sample-based monitoring using a grid of sample points of 2.8 × 2.8 km. In each point of this grid a permanent sample plot was established, where diameter and height of every tree were measured.

A very important feature of the GMS is that standing volume of a given stand is assessed from that of the above measures of tree individuals, using volume tables, and not from yield tables. In this way, the uncertainty of yield tables could be estimated by comparing sampling data and data from the yield tables. It is also possible to study the *combined* effect of sampling error of stand surveys and model error of yield tables. We carried out this, using a Monte Carlo simulation with the following steps (Figure 4):

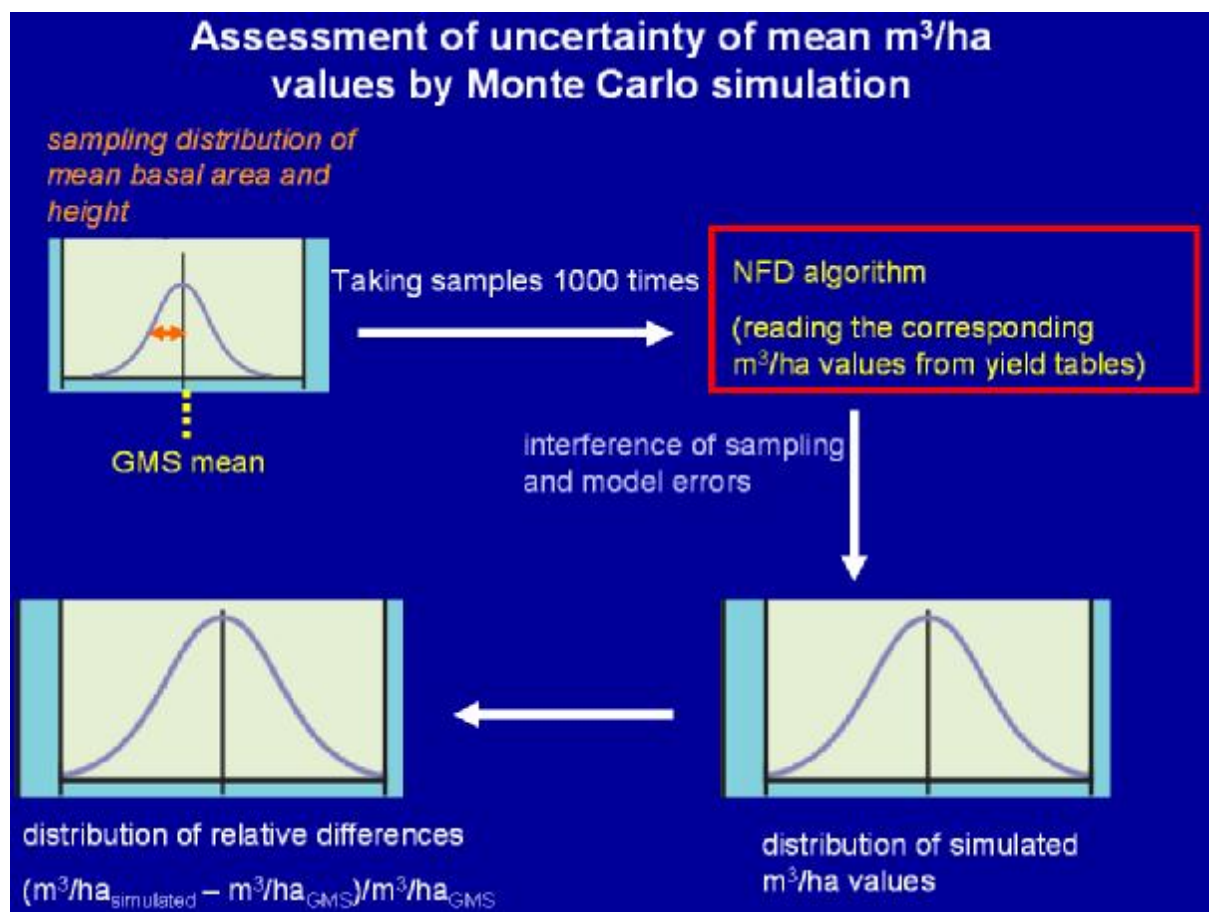


Figure 4. A schematic illustration of the Monte Carlo simulation. The input distributions are the distributions, generated by using random numbers during the simulation, of mean height and basal area from which samples were taken and used as input values of the NFD algorithm. NFD – National Forest Database; GMS – Growth Monitoring System. For further explanation see the text.

1. Calculation of average height and basal area of the dominant tree species (i.e. species of the highest proportion) for each plot from GMS data ('GMS mean' in Figure 4 represents these average values).
2. Calculation of standing volume of tree individuals from their height and diameter values using the Király 2 volume functions (Sopp and Kolozs 2000).
3. Calculation of average m^3/ha value (V_{GMS}) of the dominant tree species for each plot using plot area data.
4. Taking samples (i.e. simulated values) from distributions of mean basal area and height for 1000 times (Figure 4). These distributions were regarded as normal with means that are equal to GMS means and standard deviations that were calculated from the bootstrap distributions (i.e. that represent sampling error).
5. Calculating m^3/ha values ($V_{simulated}$) from each simulated mean height and basal area using the corresponding NFD algorithm. This algorithm reads the appropriate m^3/ha value from the yield table as a function of tree species, age, mean height, mean basal area. The later is used for correction of the m^3/ha value of the yield table in the following way:

$$V_{simulated} = (G_{simulated}/G_{yield\ table}) * V_{yield\ table},$$

where:

$V_{simulated}$ – m^3/ha value calculated from the simulated mean basal area and mean height values;

$G_{simulated}$ – simulated mean basal area;

- $G_{\text{yield table}}$ – basal area predicted by the yield table for the pure stands of the given species of the given age and height;
- $V_{\text{yield table}}$ – standing volume (m^3/ha) predicted by the yield table for the pure stands of the given species of the given age and height.
6. Calculating the relative difference between the simulated volume values with those of the GMS plots in the following way:
- $\text{RD} = (V_{\text{simulated}} - V_{\text{GMS}})/V_{\text{GMS}}$, where:
- RD – relative difference between the simulated volume (m^3/ha) value and that calculated from GMS data;
- $V_{\text{simulated}}$ – area-specific volume (m^3/ha) calculated from the simulated mean basal area and mean height values by the NFD algorithm using yield tables;
- V_{GMS} – area-specific volume (m^3/ha) calculated from GMS individual-level data.
- The obtained values gave the reference distributions used for confidence interval calculation.

Data of 12,699 GMS plots were included in the Monte Carlo simulation (Table 4). Note, that the combined effect of sampling for height and basal area, and yield table errors can be studied only if the reference distribution is based on data of several GMS plots (Figure 5). By applying relative values it became possible to use data of plots having different absolute m^3/ha values simultaneously.

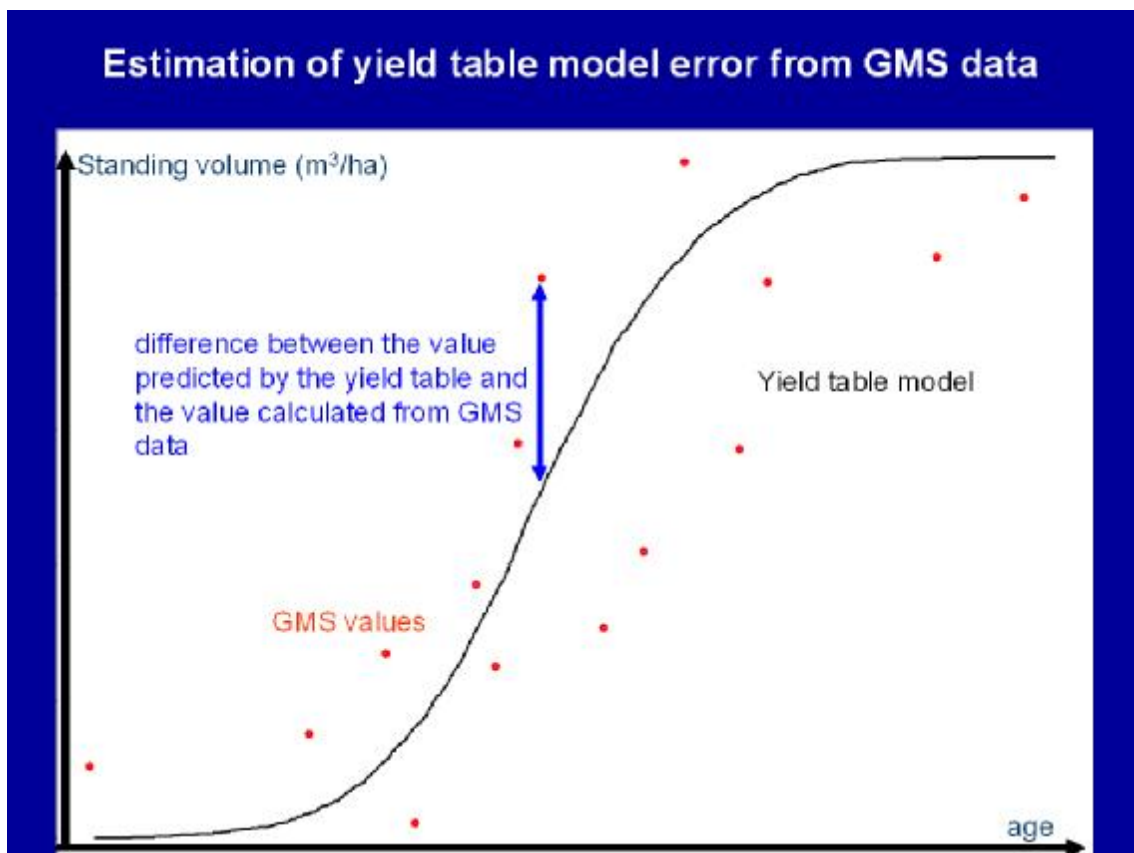


Figure 5. Schematic illustration of the assessment of errors due to the application of a yield table for a species. The prediction band of the model can be estimated from the distribution of the relative differences shown in blue in the figure.

Table 4. Sample size (i.e. no. of GMS plots) of species groups used for assessment of uncertainty of area-specific volume.

species group	sample size
Quercus robur	1233
Quercus petraea	1795
Other quercus	332
Quercus cerris	2484
Fagus silvatica	1544
Carpinus betulus	612
Robinia pseudoacacia	2261
Acer spp.	146
Ulmus spp.	10
Fraxinus spp.	442
Other hardwood species	80
Cultivated poplars	54
Indigenous poplars	141
Salix spp.	50
Alnus spp.	274
Tilia spp.	228
Other softwood species	24
Pinus silvestris	643
Pinus nigra	245
Picea abies	81
Larix decidua	17
Other conifers	3
Total	12699

Random errors of the GMS data were disregarded during the Monte Carlo simulation due to the fact, that these errors are much smaller than the modelled sampling and model errors because:

1. on a given GMS plot all trees are measured in contrast to sampling during forest planning when only a little part of the given forest subcompartment is measured, and random errors of measuring a tree individual tend to be offset by errors of measuring the other trees;
2. in the case of GMS, volume values are calculated from data of trees of the given plot whereas volume values from the yield tables were calculated from countrywide data (i.e. a national average) and consequently on plot level the former data are much more precise.

Considering these two points it can be safely concluded that random errors of GMS are negligible compared to those of forest planning sampling and yield tables. (If a considerable amount of random errors occurred in the GMS data, the estimated uncertainty of NFD volume values would be higher. This means that the applied method cannot lead to an underestimation of the real uncertainty of NFD volume values).

4. Uncertainty of NFD area-specific volume values

The results suggest a random error of +/- 20-40 % of the area-specific volume (m³/ha) values for most species groups (Figure 6), if forest planning data are gathered by the basal area sampling method. The ratio of this assessment method was 27 % in 2011. However, our

results are consistent with that of the former expert judgement (Czirok and Szabó 2011) which assumed an error of $\pm 30\%$.

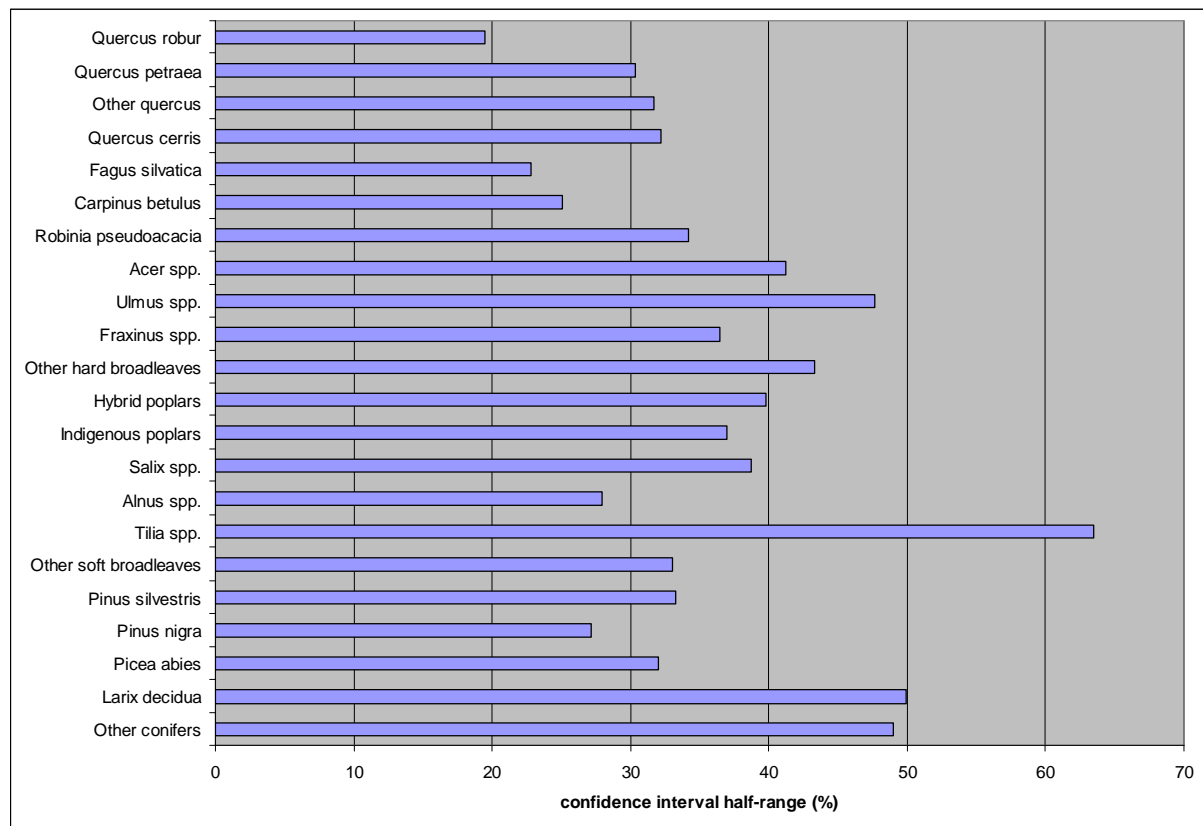


Figure 6. Percentage confidence interval half-ranges of area-specific volume (m^3/ha) values of the National Forest Database by species groups.

Some of the tree species groups had a random error higher than $\pm 40\%$. However, these represent associate tree species the proportion of which tends to be low in forest subcompartments, and the total standing volume of which was only 3 % in 2011. The average confidence interval half-range weighted by total volume of species groups was 31 % both for 2010 and 2011.

Note that the results are valid only for stands older than 20 years (black locust) and 40 years (other species) because the sampled data in section 2 above covered these age classes only. Furthermore, the results are valid only for species the proportion of which is higher than 20 percent in the given subcompartment because the analyzed GMS data were not representative for species having lower proportion. However, it can be supposed that in younger stands and in the case of species having low proportion ('rare species') random errors may be higher due to the higher variability of the stand structure and due to the patchy distribution of the trees (see also section 2). Also, younger stands and cohorts of stand with low proportion do not contribute much to either total volume or its change. Thus, in order to be conservative, a $\pm 40\%$ random error of m^3/ha data was supposed for stands older than 40 years (the total standing volume of which amounts to 63 % in 2011), and $\pm 80\%$ was supposed for younger stands. It can be safely supposed that due to their small standing volume the higher uncertainty of rare species does not increase the overall average uncertainty more than 10 percent, so an overall random error of $\pm 40\%$ can be regarded as a conservative estimation for older stands. Considering younger stands, it must be noted that the more or less systematic basal area sampling method is very rarely applied in them. Instead, a highly preferential

sampling method (forecasting using yield table, see section 2) is used which may lead to a higher uncertainty which has not been estimated yet by statistical methods. That is the reason why a much higher uncertainty value was used for these stands.

5. Uncertainty of forest subcompartment areas

The error of subcompartment areas can be modelled from prescriptions related to mapping precision. That is, a maximum error of 6 m, relative to the “true” location, of any point of the border of the subcompartment has been allowed until recently. It means that the difference between the true and the mapped location of the border lines of a given subcompartment can be maximum 6 m (Figure 7). It must be emphasized that this maximum distance is clearly an overestimation of the real precision of mapping practice according to expert judgments. Still, an error of 6 m was supposed, when calculating area uncertainty for the sake of conservativeness.

According to a case study carried out by an expert (Mezei 2011) modelling using rectangles to approximate the shape of the subcompartments gives reliable results that do not differ considerably from those when the real shapes are used for the calculation of the possible highest differences between the true and the mapped areas (Figure 7). Thus, due to practical reasons, the above-mentioned assumption of rectangular shape was applied in this analysis, too.

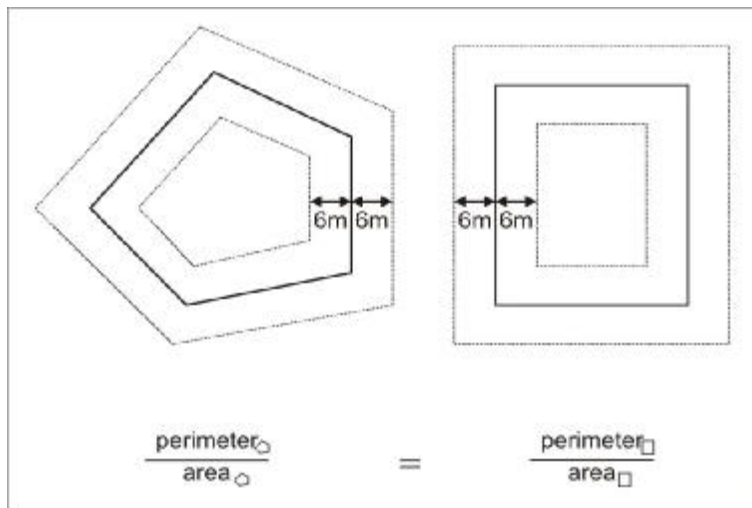


Figure 7. Schematic representation of errors of +/- 6 m in the location of the border lines.

Assuming thus that all subcompartments are rectangular, and that the error of the location of their borderlines is random, we calculated the maximum positive and negative difference between the size of the true and the mapped areas (Figure 8). Note, that application of range instead of confidence interval is a clear overestimation of random error, too. However, as the IPCC 2006 GL suggests on page 3.21 that when the only available information about uncertainty is the range it can be treated as a confidence interval (with 95 % confidence level). Thus, the percentage differences between the possible maximum and the possible minimum size of the mapped areas were applied as uncertainty values. Using Monte Carlo simulation, the error of the total area of forested land is 0.03 % on country-level even by this highly conservative approach.

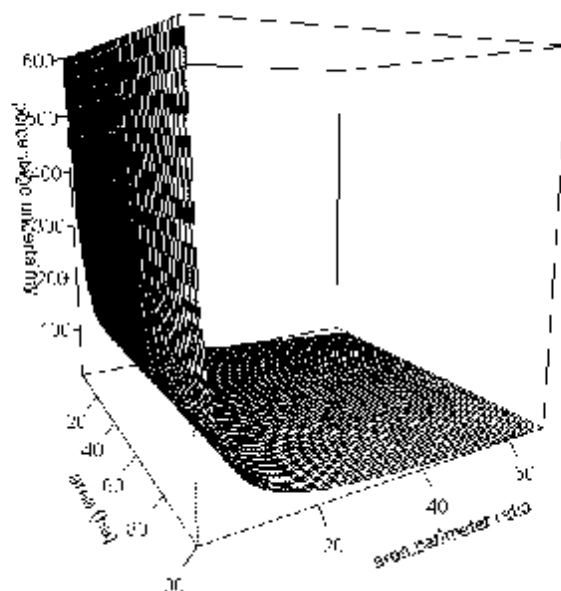


Figure 8. Percentage uncertainty of the area of forest subcompartments as a function of the size of the area and the area:perimeter ratio assuming a 6 m error in the location of the border lines.

6. Uncertainty of standing volume by species groups on country level

The uncertainty of standing volume of species groups on country level was studied by a Monte Carlo simulation. These volume values are multiplied by the corresponding emission factor values (carbon fraction, wood density, root-to-shoot ratio; see NIR Chapter 7.3.1.2.1). Thus, their uncertainties are essential for the uncertainty analysis introduced in NIR Chapter 11.3.1.5).

The main steps of this Monte Carlo simulation were the following:

1. Taking samples of distributions of mean subcompartment areas and area-specific volumes on forest subcompartment level for 1000 times. Both of the area and the area-specific volume distributions were regarded as normal distributions. Standard deviations of area distributions were calculated from the range rule, according to which *standard deviation* = $(Maximum - Minimum)/4$ (a detailed description of range rule is available on: <http://statistics.about.com/od/Descriptive-Statistics/a/Range-Rule-For-Standard-Deviation.htm>). The range was calculated as described above. For the area-specific volume distributions, standard deviations were calculated in a similar manner, i.e. the estimated confidence interval half-range was divided by 2 (see the same webpage cited above). The means of the distributions were the corresponding means stored in the NFD. In this way, area and area-specific volume samples were taken 1000 times from ca. 2 x 1.1 million distributions (which equals the number of forest subcompartment – species combinations).
2. Multiplying the corresponding area values by the corresponding area-specific volumes. These multiplications gave standing volume values on forest subcompartment level that provided distributions for the uncertainty estimation.
3. Summing up the subcompartment level standing volume values by species groups and calculating the percentage confidence interval of the means on country level. Table 5

shows the results for two KP categories (i.e., Forest Management, or FM, and Afforestation and Reforestation, or AR) separately.

Table 5. Standing volume and stock change uncertainties of species groups. sd – standard deviation of the mean.

Species groups	FM								AR							
	standing volume 2011				stock change 2010-2011				standing volume 2011				stock change 2010-2011			
	mean 10 ³ m ³	sd %	CI upper %	CI lower %	mean 10 ³ m ³	sd %	CI upper %	CI lower %	mean 10 ³ m ³	sd %	CI upper %	CI lower %	mean 10 ³ m ³	sd %	CI upper %	CI lower %
Quercus robur	32413.1	0.2	0.4	-0.3	250.4	30.2	58.9	-59.4	936.0	1.1	2.1	-2.1	120.1	12.6	24.2	-25.0
Quercus petraea	44975.0	0.1	0.3	-0.3	-212.7	-44.8	85.6	-88.5	132.1	2.2	4.1	-4.4	19.5	26.0	51.1	-49.5
Other quercus	6001.7	0.4	0.8	-0.8	193.3	18.4	35.3	-35.4	70.9	3.3	6.4	-6.3	8.8	39.6	81.1	-74.5
Quercus cerris	45488.3	0.1	0.3	-0.3	341.5	23.6	46.8	-43.8	109.1	2.1	4.2	-4.1	16.9	20.5	43.4	-38.3
Fagus sylvatica	39328.1	0.2	0.4	-0.4	-73.4	-142.4	282.9	-283.9	4.7	6.7	13.4	-13.2	0.6	-133.3	262.6	-261.3
Carpinus betulus	17326.6	0.1	0.3	-0.3	40.1	76.3	140.1	-149.0	14.7	2.6	5.1	-5.2	2.0	25.7	50.1	-48.2
Robinia pseudoacacia	43681.2	0.2	0.4	-0.3	123.9	42.0	77.0	-83.5	5156.5	0.6	1.1	-1.1	495.5	8.1	16.3	-16.6
Acer sp.	3650.5	0.3	0.6	-0.7	191.5	7.9	14.9	-16.0	35.3	3.5	6.7	-6.8	4.4	40.3	71.8	-81.0
Ulmus sp.	588.6	0.7	1.5	-1.4	17.2	28.4	56.9	-53.7	91.9	3.6	6.7	-6.8	8.4	76.4	148.8	-146.9
Fraxinus sp.	12000.0	0.2	0.4	-0.4	278.6	12.6	24.6	-23.6	127.5	1.7	3.3	-3.3	17.8	17.0	33.1	-33.3
Other hard broadleaves	3303.8	0.4	0.8	-0.8	17.4	67.1	134.5	-130.0	91.2	2.9	5.6	-5.9	10.8	40.7	76.5	-80.1
Hybrid poplars	11852.0	0.4	0.8	-0.7	-168.6	-53.2	109.9	-101.3	4086.6	0.7	1.4	-1.4	250.8	18.4	36.3	-36.0
Indigenous poplars	11750.5	0.3	0.6	-0.7	215.6	19.4	37.8	-37.4	1277.0	1.0	2.0	-1.9	146.7	14.6	29.6	-27.9
Salix sp.	4803.2	0.5	1.1	-1.1	23.3	79.3	155.0	-147.6	70.5	4.9	10.0	-9.8	5.3	131.7	260.0	-249.4
Alnus sp.	10375.9	0.3	0.6	-0.6	125.9	27.3	53.6	-52.1	89.7	2.6	5.1	-5.4	6.9	42.8	88.2	-82.2
Tilia sp.	6698.7	0.3	0.6	-0.7	60.8	47.2	97.1	-92.7	3.3	4.9	9.9	-9.4	0.6	26.6	52.3	-52.7
Other soft broadleaves	1399.8	0.7	1.5	-1.4	41.3	31.9	60.1	-60.9	8.8	12.1	23.8	-23.9	0.8	173.8	335.3	-357.6
Pinus silvestris	34818.2	0.2	0.4	-0.4	85.8	90.5	176.9	-178.7	175.3	3.1	6.0	-6.0	10.8	66.4	127.7	-129.4
Pinus nigra	11590.5	0.3	0.6	-0.6	-32.5	-146.5	282.4	-294.3	307.6	2.1	4.1	-4.1	19.3	31.8	64.7	-58.4
Picea abies	5656.4	0.6	1.1	-1.1	35.3	128.9	261.1	-247.9	31.5	5.3	10.1	-10.8	2.6	61.6	124.6	-118.4
Larix decidua	1290.7	0.7	1.3	-1.3	20.5	55.3	103.1	-103.9	5.7	7.1	13.4	-14.2	0.6	68.8	132.0	-139.9
Other conifers	373.7	1.3	2.7	-2.5	12.9	46.7	89.4	-88.2	1.9	10.8	19.9	-21.8	0.2	64.5	125.9	-126.1

7. References

- Czirok, I., Szabó, Sz. 2011: Expert judgement, reference no.: 2011/1.
- Davison, A.C., Hinkley, D.V. 1997: Bootstrap Methods and Their Application, Cambridge University Press.
- Mezei, L. 2011: Expert judgement, reference no.: 2011/3.
- Sopp, L., Kolozs., L. 2000: Fatömegszámítási táblázatok. [Volume tables.] Állami Erdészeti Szolgálat, Budapest.