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

SUPPLEMENT TO THE NATIONAL GREENHOUSE GAS INVENTORY OF HUNGARY

by: Tamás Tobisch
Department of Forestry
National Land Center

Budapest, 2022

Table of contents

1. Foreword

The chapters of this document are modified versions of the corresponding chapters of Tobisch and Kottek (2013) because the original calculation methods were developed in order to assess uncertainty related to forest biomass in a more sophisticated way. The main difference between the original and the current methods is that instead of applying constant uncertainty values for m^3/ha statistics of the Hungarian National Forest Database (NFD) for younger and older stands, separately (see the linked document above), confidence interval half-ranges of various species groups were modelled as a function of the mean.

2. Introduction

The carbon content of forest biomass pool is assessed from NFD growing stock data. In the NFD, standing volume data by tree species are stored on forest sub-compartment 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 here). 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.

3. 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. It should be noted that for height measurements, which is needed as inputs for the yield tables, trees are chosen preferentially. Consequently, applying the calculated *sampling error* of average height may lead to an underestimation the true (accurate) uncertainty of the volume data. However, when combining sampling and modelling errors by comparisons simulated and real data (see Chapter 4), this underestimation is eliminated since the calculated combined errors involve *all* differences between the simulated and real data.

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) 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).

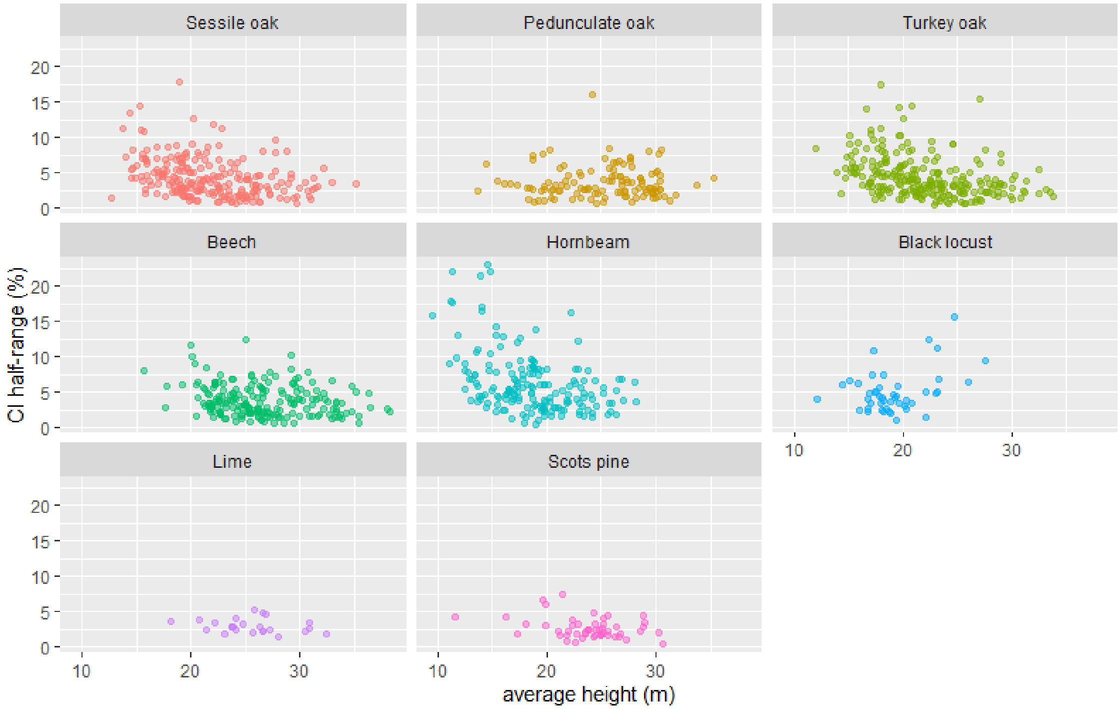

Figure 1 95 % confidence interval half-range over the mean height by species groups.

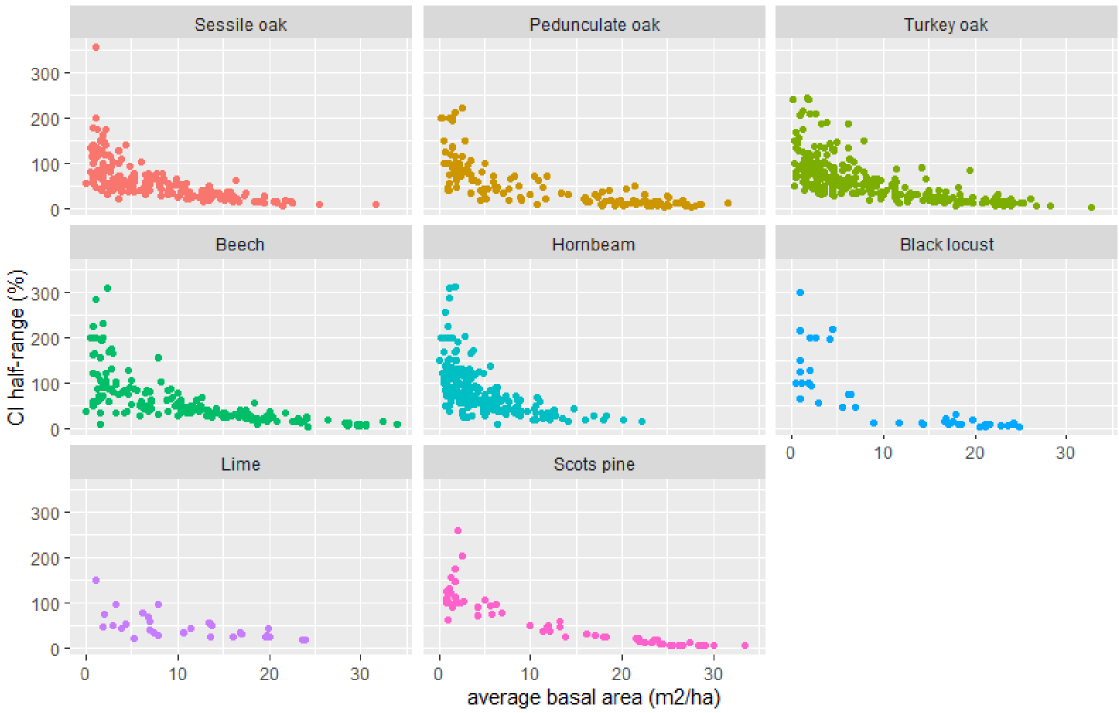

Figure 2 95 % confidence interval half-range over the mean basal area by species groups.

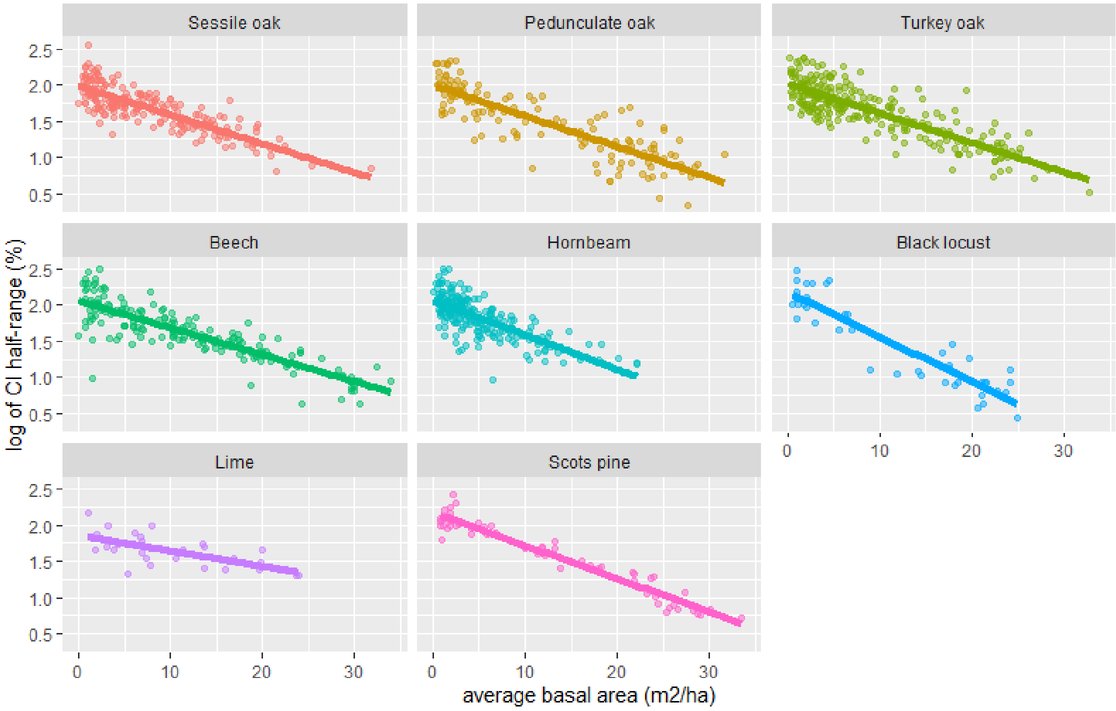

Figure 3 Linear (logarithmic) regression of 95 % confidence interval half-range over the mean basal area by species groups.

According to the results of the bootstrap analysis, the uncertainty of average height is independent from the mean and within +/- 10 % in most cases (Figure 1). This result is consistent with a previous 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 +/- 10 % was assumed throughout our investigation, independently from the average height or tree species.

The uncertainty of average basal area strongly depends on its absolute value (Figure 2 and Figure 3). In the case of species, the proportion of which is low in the given sub-compartment, percentage confidence interval of basal area can be extremely high, 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 percentage confidence interval half-range for those species groups for which we received sampling data enough for quantifying uncertainties (Figure 3). For those species groups for which we did not receive enough data (i.e., sampling size was lower than 500) we applied an 'average model' which was based on all of the received data regardless of species.

However, in some cases, the linear regression models underestimate the true uncertainty for small values of average basal area. In order to avoid such underestimation, we used the maximum of the calculated uncertainty values of the given species group for basal areas smaller than 1 m^2/ha.

4. 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 here.

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):

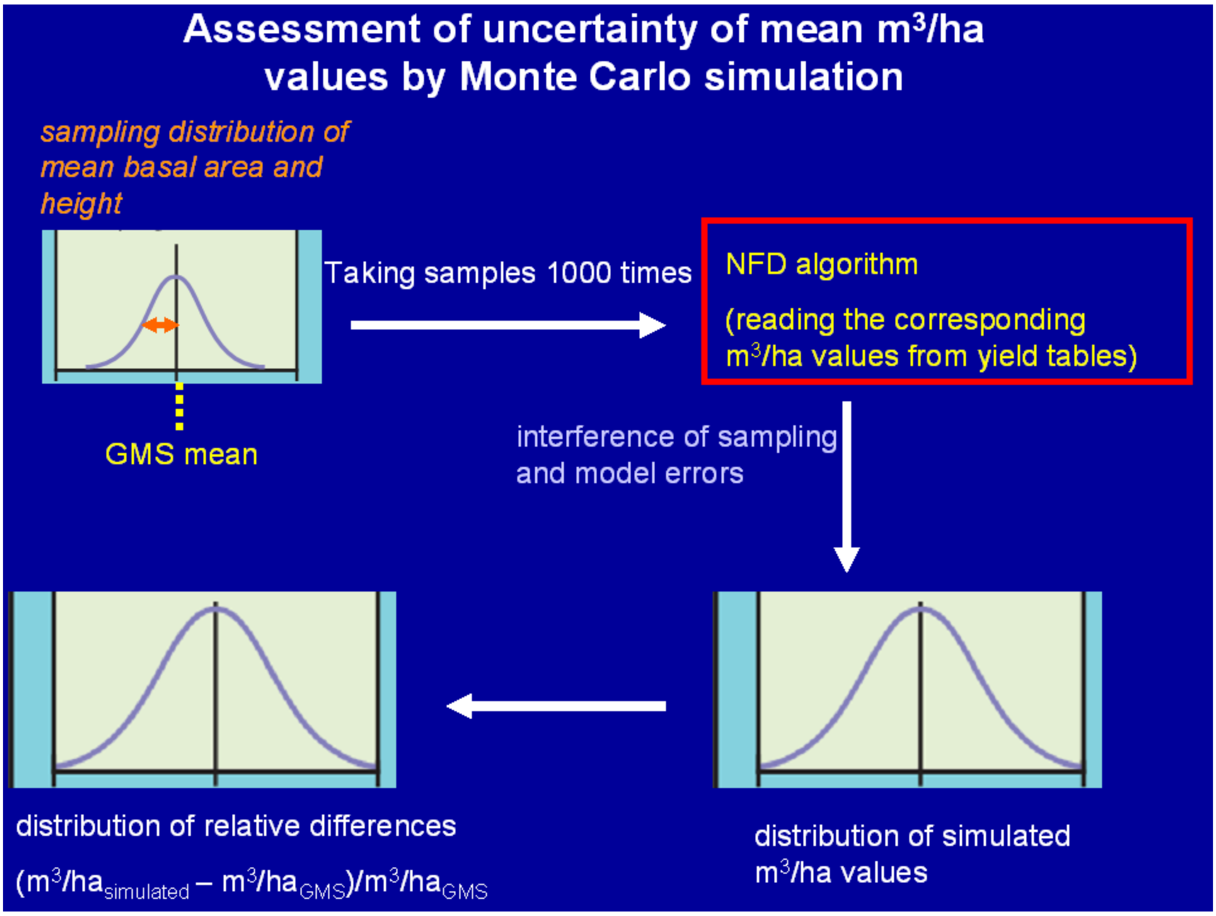

Figure 4 A schematic illustration of the Monte Carlo simulation. The input distributions are the distributions, generated by using random numbers during the simulation, of mean height and basal area from which samples were taken and used as input values of the NFD algorithm. NFD – National Forest Database;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

   $G_{yield\ table}$ – basal area predicted by the yield table for the pure stands of the given species of the given age and height;

   $V_{yield\ table}$ – 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:

   $RD = (V_{simulated} – V_{GMS})/V_{GMS}$, where:

   RD – relative difference between the simulated volume (m^3/ha) value and that calculated from GMS data;

   $V_{simulated}$ – 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. 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). Thus, ‘classes’ were defined based on the m^3/ha values of GMS plots, and the combined errors were calculated for these classes.

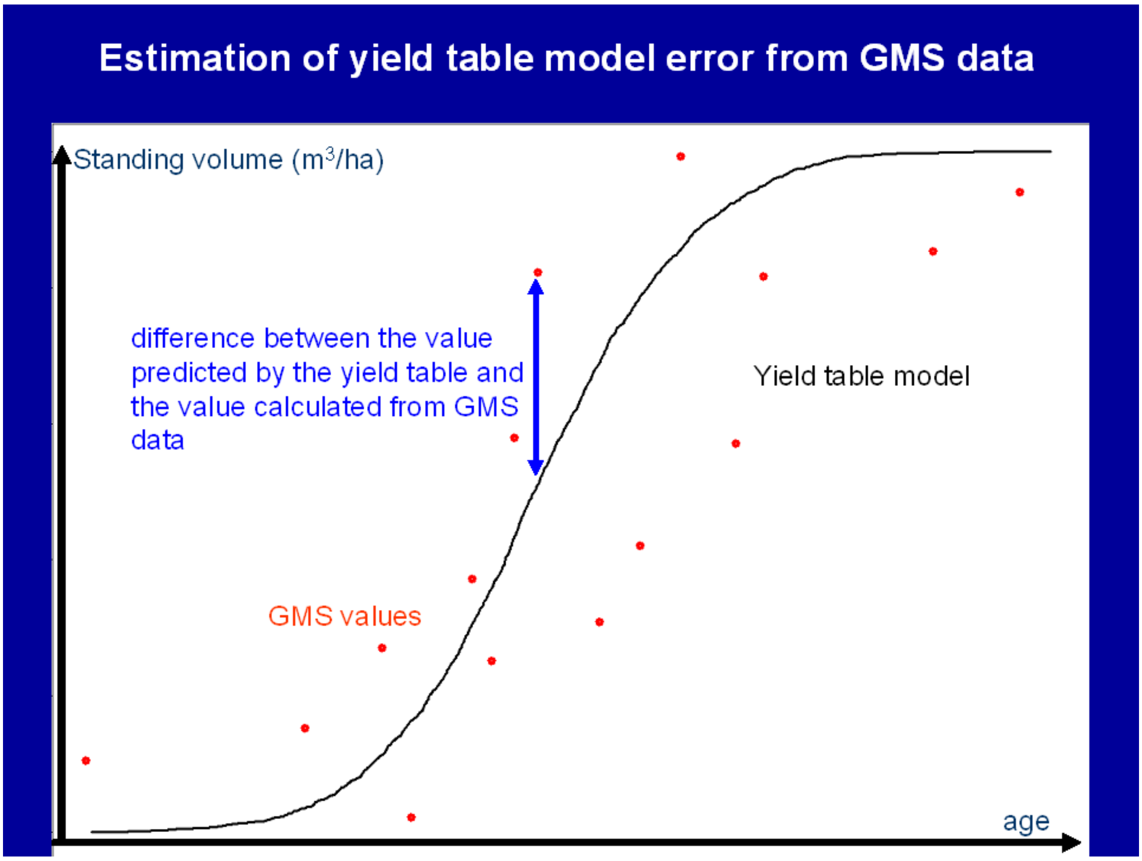

Figure 5 Schematic illustration of the assessment of errors due to the application of a yield table for a species. The prediction band of the model can be estimated from the distribution of the relative differences shown in blue in the figure.

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).

The resulting combined errors together with the sampling errors are shown in Figure 6.

Figure 6 Sampling and combined (total) error of average m³/ha values by species groups.

5. Uncertainty of NFD area-specific volume values

The results suggest a random error of +/- 20-40 % of the area-specific volume (m^3/ha) values for most species groups (Figure 6). This is consistent with a former expert judgement (Czirok and Szabó 2011) that assumed an error of +/- 30 %.

Just as in the case of the sampling error of basal area, the combined error of mean m^3/ha values can be modelled as a function of the mean (Figure 7). For those species, for which there were not enough data (at least four m^3/ha classes of at least 50 plots), we applied the 'average model' (which was calculated from all data).

In the NFD, for a given year (so called 'closed statistical state' of a year) we have approximately 1.3 million records of various species of approximately 0.5 million forest sub-compartments. In each record, some mean statistics (height, basal area, area-specific volume, closure etc.) are stored. For assessing area-specific growing stock uncertainty for the GHG inventory, confidence interval half-ranges are modelled from the models described above with the restriction that the modelled value must be neither higher than the modelled maximum value nor lower than the modelled minimum value. In other words: for extrapolation of the models the 'observed' minimum and maximum uncertainty values are applied.

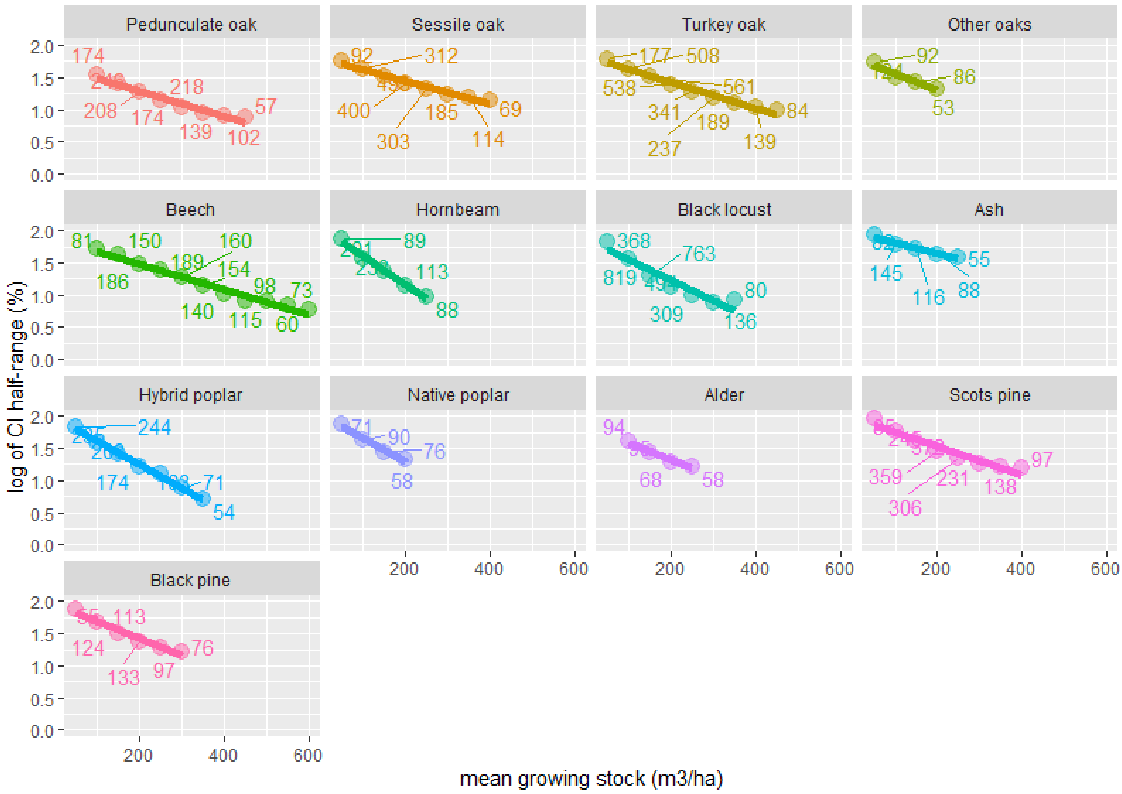

Figure 7 Percentage confidence interval half-ranges of area-specific volume (m^3/ha) values modelled by the mean by species groups. The numbers indicate the number of GMS plots representing the given point.

6. 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.

Sopp, L., Kolozs., L. 2000: Fatömegszámítási táblázatok. [Volume tables.] Állami Erdészeti Szolgálat, Budapest.