A review of estimation and processing procedures of National Forest Inventory

July 3, 2019

Contents

1 List of terms and abbreviations

s .... area of a single plot

CSV .... comma separated values

Characteristic categorical value associated with population element or strata

DBH .... diameter at breast height (1.3 m)

Domain ..... sub-population, part of population distinguished accordingly to the factors values

Factor .... cathegorical characteristic of tree or stand, that is used for subdivision of total results

IID .... independent and identically distributed

Parameter .. population parameter is function of variables of study

Population .. forest population is a set of finite number of trees in a given forest area

Record ...... single entry in the sample

Stand .... part of forest that differ by combination of characteristics (a kind of forest strata)

Stratification division of (forest) area into disjoint sub-areas (strata)

Tract .... a group of 4 plots

Variable ..... numerical value associated with population element or strata

2 National Forest Inventory

National Forest Inventory (NFI) is a system of annually sample surveys of country forest lands.

2.1 Forest definition

The definition of forest used to NFI:

1. minimum area of forest parcel: 0.1 ha;
2. minimum crown coverage (or equivalent stocking): 30 percent;
3. minimum height of trees at age of maturity: 5 m;
4. minimum forest width: 20 m.

2.2 Sampling frame and sample units

The random selection of trees from the population is replaced by the (random) selection of sampling units from a sampling frame. A grid of squares sized 5.0 by 5.0 km is imposed on the map of country territory. Every inventory square contains a cluster of four inventory sample plots, considered as an inventory tract. The tract is randomly placed within an inventory square. Each inventory tract and sample plot are assigned to one of five temporal panels as is illustrated at figure 1. Proposed sampling frame corresponds to the intensity of annual sampling of 20 percent from the total amount of tracts, given the length of inventory cycle in 5 years. Also, yearly inventories could be considered as separate samples.

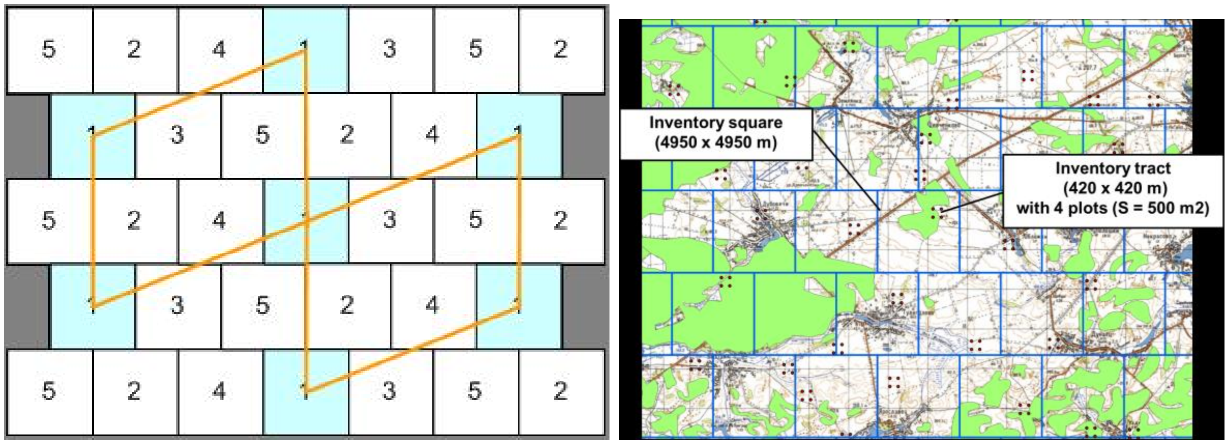


Figure 1: NFI sampling frame

2.3 Calculation units

Outlined sampling frame defines the stratified cluster random sample with square grid cells as separate strata, inventory tracts as clusters, and inventory plots as sampling units. Despite the distance between sample plots are accepted so they shouldn't fall into the same forest, some characteristics frequently correlate for the tract level (species, forest type, age, disturbances etc.) and considering them as independent observations could lead to biased estimates. Thus the basic calculation unit for estimations of NFI parameters is selected to be the tract.

2.4 Sample plot

NFI uses the fixed circular plot sampling. A sampling plot consists of set of circular plots (nested plots) for measurements of trees of different diameters as presented at figure 2 and table 2.

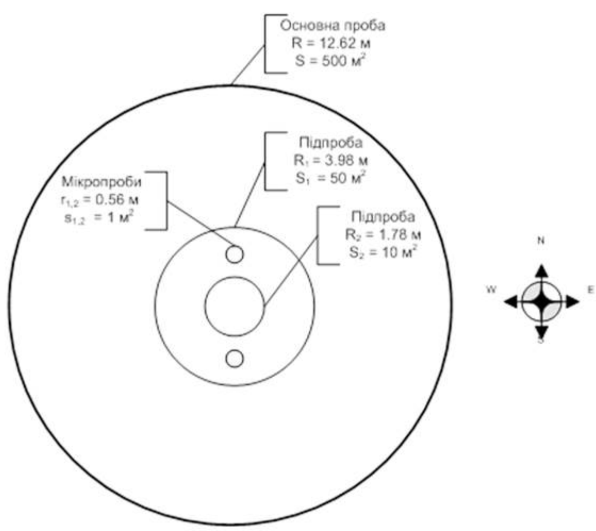


Figure 2: Sample plot design

Under fixed area plot sampling the radii define the tree sampling function. The tree sampling function is not a stochastic process, but simply an association rule between a sampling point (plot center) and the number of population elements (trees). However, as we assume that location of sampling point is random outcome of stochastic process generating points over population area, the selection of trees is a random process [2].

For the calculation, the tree variables are weighted to the plot area by multiplier defined as $M_s = A_0/A_s$, $s = 0, \ldots, 3$, where A_s is area of nested plot. Currently, micro plots are not used for processing.

Plot	Radius, m	Area, m^2	Condition	Multiplier, M_s
(Main) Plot 0	12.62	500	DBH $\geq$ 14	1
Plot 1	3.98	50	DBH $\in$ [6, 14)	10
Plot 2	1.78	10	DBH $\in$ [2, 6)	50
(Micro) Plots 3	0.56	1	DBH $\leq$ 2 or H $\leq$ 1.3	0

Table 1: Tree measurement conditions on sample plot

All forest plots are investigated for data collection by field teams.

At field survey the area of each sample plot also shall be divided into segments (subplots) related to the different categories of forest (non-forest) lands. Thus the boundaries of different forest stands are mapped within the plot. The trees are attributed to different domains compiled from the subplots with identical characteristics.

List of variables and characteristics to be collected on the sample plots is designated by the Instruction on the conducting of NFI [].

3 Data storage

NFI geo-database is created as Field-Map project [https://www.fieldmap.cz/] that combines the sets of *db*-tables and *shp*-layers (figure 3).

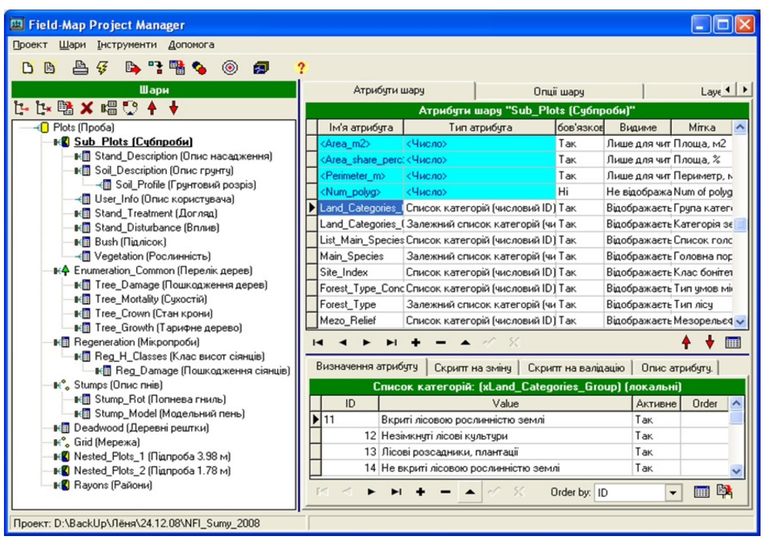


Figure 3: NFI Field-Map pproject

The database includes the groups of characteristics for different objects:

1. plot: GPS coordinates, administrative information, relief;
2. segments (subplots): user info, description of forest stand (layers, treatment, disturbance, forest type, site index), undergrowth and bushes, herb vegetation, deadwood, soil;

and variables for:

1. tally trees: location, species, measurements (DBH); state (mortality, damages), estimations (stem quality, Kraft classes) etc.;

2. model (or tariff) trees: measurements (H, age), estimations (crown state);
3. regeneration trees: measurements (H), state (damages);
4. stumps: diameter, state (decays) etc.

Annual databases are stored separately but combined for the common estimations for several years or inventory circle.

The structure of database tables is presented in Annex 1 to this document.

4 Data modeling

From many variables observed for each plot tree, only species type and the diameter (DBH) are available for all trees in sample. The tree height (H) is measured for the sample of second stage (model) trees, and no measurements for tree volumes. These values should be estimated from available measurements. We apply two models for this: for height and for volume. They are based on volume assortments tables for standing trees [3]. These tables provide data of growth level, DBH, Height and Volume for 14 kinds of species.

4.1 Height modeling

Height is estimated from DBH with the following function

$$H = (a + b * G) * arctan(c * DBH), \tag{1}$$

where a, b and c are parameters fitted for each species separately. For example, for one of species the data and fitted curves are presented at figure 4. Estimated parameter values are provided at table 2.

parameter	a	b	c	d	e
value	0.4	0.6	0.2	22	0.4

Table 2: Fitted parameter values for species kind 2

To apply the model we have to estimate growth level of each stand. We define separate stand by combination of IDPlots, IDSub_Plots and Species. For all model trees we can compute growth level by formula

$$G = \left(\frac{H}{arctan(c * DBH)} - a \right) \Big/ b.$$

If some stand has several model trees we can average their values to get growth level of this stand. If stand doesn't have model trees, we use computed growth level from stand description.

In case if we still cannot achieve growth level, we assign it to the minimum value for current Sub_Plot.

Using growth level of stand we can use equation 1 to compute heights of all trees.

This process is illustrated at figure 5 with 17 trees and 7 model trees.

In the inventory of Sumy region in 2008 the relation between model height and predicted height is illustrated at figure 5.

4.2 Volume modeling

Tree volume is represented by the equation

$$V = f * \pi \left(\frac{DBH}{200} \right)^2 * H, \tag{2}$$

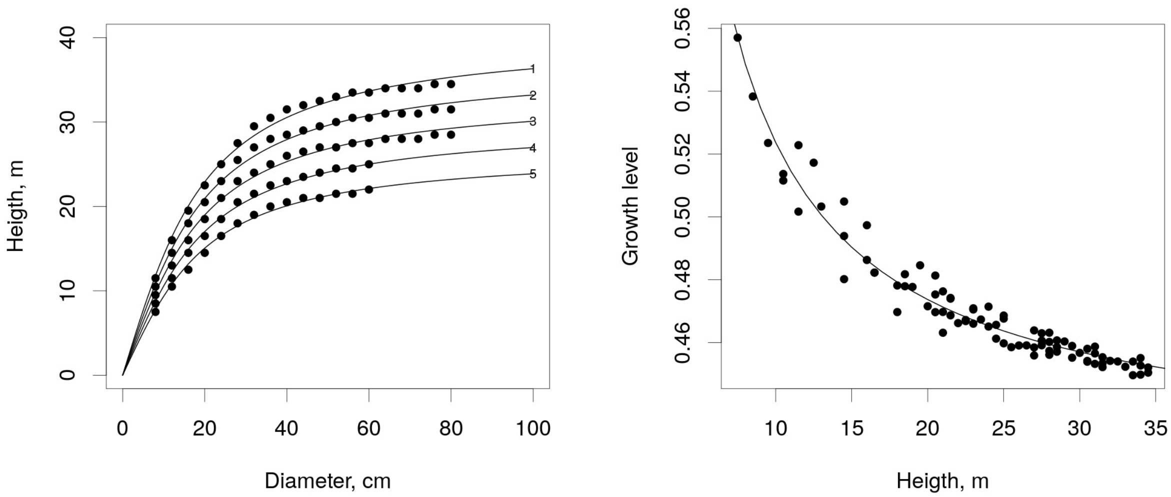


Figure 4: Height and form factor models for species kind 2

where f is form factor, that is characterise tree shape and is a ratio between cylinder and tree volume.

Using volume assortments tables we can compute form factor as

$$f = \frac{V}{\pi(DBH/200)^2 * H}. \tag{3}$$

To predict form factor the following formula is used

$$f = d + \frac{e}{H},$$

where d and e are parameters fitted for each species separately. For species kind 2, data and fitted curve are presented at figure 4. Estimated parameter values are provided at table 2.

5 Statistical estimation

Having conducted an inventory, we need a way to estimate total value of specific characteristic X and its accuracy for the whole population with total area S. Let our sample consists of N tracts with area s each, generally $4\pi r^2$. Each tract has N_i, $i = 1, \ldots, N$ records. The values of our characteristic are X_{ij}, $j = 1, \ldots, N_i$. Then, let us denote total value for tract i to be

$$X_i = \sum_{j=0}^{N_i} X_{ij}, \ i = 1, \ldots, N.$$

5.1 Total value and it's accuracy

Usual unbiased estimates of mean value and variance are

$$\bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i,$$

$$S_X^2 = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \bar{X})^2.$$

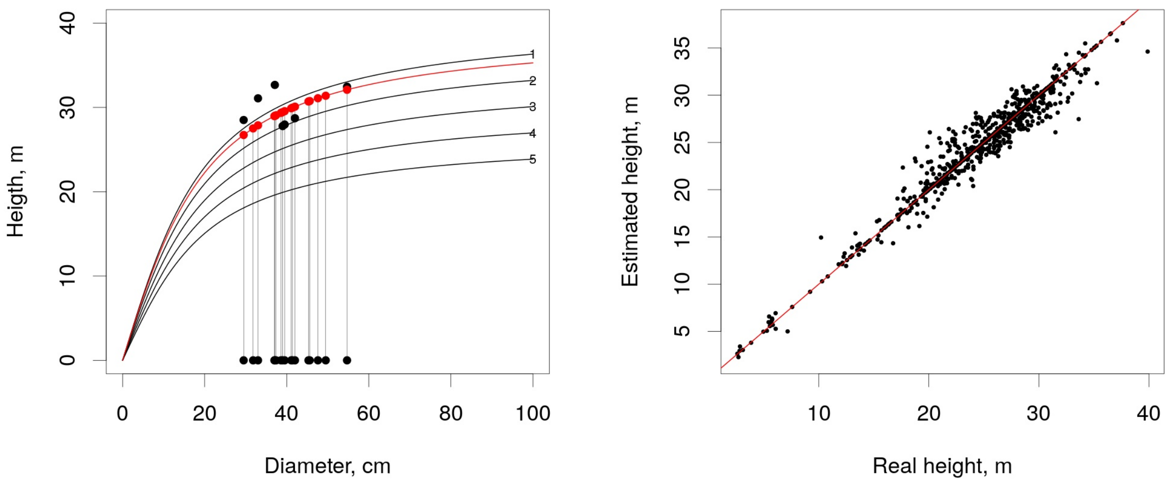


Figure 5: Height estimation

The estimator for the total value is

$$\hat{X} = \frac{S}{s}\bar{X}. \tag{4}$$

Since $S^2_{\bar{X}} = S^2_X/N$, the estimator for the variance of total value is

$$S^2_{\hat{X}} = \frac{S^2}{s^2}S^2_{\bar{X}} = \frac{S^2}{s^2 N}S^2_X.$$

To estimate accuracy we will use coefficient of variation (CV). It is defined as

$$\mathbf{CV}\, X = \frac{\sqrt{\mathbf{Var}\, X}}{\mathbf{E}\, X}$$

and estimated by

$$\widehat{\mathbf{CV}}\, \hat{X} = \widehat{\mathbf{CV}}\, \bar{X} = \frac{S_X}{\bar{X}\sqrt{N}}.$$

Since the number N is usually quite large (>1000) we can approximate distribution of $\hat{X}$ by $N(\hat{X}, S^2_{\hat{X}})$. Then, we can select confidence level $0.5 < \alpha < 1$. Let us denote $C_\alpha = Q^{N(0,1)}((1+\alpha)/2)$ the quantile of standard normal distribution of level $(1+\alpha)/2$. Using it we can construct confidence interval

$$[\hat{X}(1 - C_\alpha \widehat{\mathbf{CV}}\, \hat{X}), \hat{X}(1 + C_\alpha \widehat{\mathbf{CV}}\, \hat{X})].$$

Usually $\alpha = 0.95$ is used, then $C_\alpha \approx 1.96$.

5.2 Estimation of ratios

Let we have two characteristic X and Y. Our goal is to estimate their ratio $R = X/Y$. Let us denote normalized variables $X_n = X/\mathbf{E}\, X$, $Y_n = Y/\mathbf{E}\, Y$ and $R_n = X_n/Y_n$. At it was shown in [1], first order approximations of mean and variance are:

$$\mathbf{E}\, R_n \approx \frac{\mathbf{E}\, X_n}{\mathbf{E}\, Y_n} = 1,$$

$$\mathbf{Var}\, R_n \approx \mathbf{Var}\, X_n - 2\,\mathbf{Cov}(X_n, Y_n) + \mathbf{Var}\, Y_n.$$

Thus, estimates for the unnormalized variable are

$$\mathbf{E}\,R \approx \mathbf{E}\,R_n \frac{\mathbf{E}\,X}{\mathbf{E}\,Y} = \frac{\mathbf{E}\,X}{\mathbf{E}\,Y},$$

$$\begin{aligned}\mathbf{CV}\,R &= \frac{\sqrt{\mathbf{Var}\,R}}{\mathbf{E}\,R} = \sqrt{\mathbf{Var}\left(\frac{R}{\mathbf{E}\,R}\right)} = \sqrt{\mathbf{Var}\left(\frac{R_n}{\mathbf{E}\,R_n}\right)} \\ &\approx \sqrt{\mathbf{Var}\,R_n} \approx \sqrt{\mathbf{Var}\,X_n - 2\,\mathbf{Cov}(X_n, Y_n) + \mathbf{Var}\,Y_n} \\ &= \sqrt{\frac{\mathbf{Var}\,X}{(\mathbf{E}\,X)^2} - 2\frac{\mathbf{Cov}(X,Y)}{\mathbf{E}\,X\,\mathbf{E}\,Y} + \frac{\mathbf{Var}\,Y}{(\mathbf{E}\,Y)^2}}.\end{aligned}$$

Let for X and Y we observe their tract values X_i, Y_i, $i = 1, \ldots, N$. Then we can estimate the values in the previous formulae accordingly to the next table:

Value	Estimate	Formula
$\mathbf{E}\,X$	$\bar{X}$	$\frac{1}{N}\sum_{i=1}^{N} X_i$
$\mathbf{E}\,Y$	$\bar{Y}$	$\frac{1}{N}\sum_{i=1}^{N} Y_i$
$\mathbf{Var}\,X$	S_X^2	$\frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2$
$\mathbf{Var}\,Y$	S_Y^2	$\frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^2$
$\mathbf{Cov}(X,Y)$	S_{XY}^2	$\frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})$

Sample estimates of mean and coefficient of variation for R are:

$$\bar{R} \approx \frac{\bar{X}}{\bar{Y}}, \tag{5}$$

$$\widehat{\mathbf{CV}}\,R \approx \sqrt{\frac{S_{\bar{X}}^2}{\bar{X}^2} - 2\frac{S_{\bar{X}\bar{Y}}^2}{\bar{X}\bar{Y}} + \frac{S_{\bar{Y}}^2}{\bar{Y}^2}} = \sqrt{\frac{1}{N}\left(\frac{S_X^2}{\bar{X}^2} - 2\frac{S_{XY}^2}{\bar{X}\bar{Y}} + \frac{S_Y^2}{\bar{Y}^2}\right)}. \tag{6}$$

5.3 Missing values

During the data collection, some variables values can be missed or assigned incorrect values for the different reasons. Later these may lead to incorrect or biased estimates. Ideally, the mistake values should be identified and corrected during the field control. Established control procedures should check the consistency and completeness of the data. To deal with missing values, several approaches can be used.

One of the simplest and popular method is mean value imputation. In that method records with missing values X_{ij} are replaced by mean values over sample. Denote missing value by NA.

If we suppose that values are missed completely at random (MCAR), then unbiased estimate for mean is

$$\bar{X}^M = \frac{1}{N - N_M}\sum_{i=1}^{N}\sum_{X_{ij}\neq \mathrm{NA}} X_{ij},$$

where N_M is the total number of missing values

$$N_M = \sum_{i=1}^{N}\sum_{j=1}^{N_i} \mathbf{1}_{\{X_{ij}=\mathrm{NA}\}}.$$

Imputation of these values instead of NA provides corrected values X_{ij}^M. Then we can compute total values in tracts as

$$X_i^M = \sum_{j=1}^{N_i} X_{ij}^M.$$

These values can be used to compute sample variance

$$S_X^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i^M - \bar{X})^2.$$

5.4 Usage of indicators

Often, we are interested in a characteristic X in the part of the total population (called a domain) defined by conditional indicator I. It's sample values are I_{ij}, $i = 1, \ldots, N$, $j = 1, \ldots, N_i$. If we define new characteristic $X_{ij}^I = X_{ij} I_{ij}$, then total tract values for the domain will be

$$X_i^I = \sum_{j=0}^{N_i} X_{ij}^I,\ i = 1, \ldots, N.$$

Similarly, results (4) can be applied to estimate average and total values and CV:

$$\bar{X}^I = \frac{1}{N}\sum_{i=1}^{N} X_i^I, \tag{7}$$

$$S_{\bar{X}^I}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i^I - \bar{X}^I)^2, \tag{8}$$

$$\hat{X}^I = \frac{S}{s}\bar{X}^I, \tag{9}$$

$$\widehat{\mathbf{CV}}\ \hat{X}^I = \frac{S_{\bar{X}^I}}{\bar{X}^I}. \tag{10}$$

For example, forested land indicator can be used to estimate total forested area.

Usually the tables constructed contain estimates not only for the whole population, but different estimates for values of a factor. In general, several factors can be used for different domains. Let we have K-factor table ($K = 1$, 2 or 3). Let the factors values are $F_{ij}^k \in \left\{C_1^k, C_2^k, \ldots, C_{N_k}^k\right\}$, $k = 0, \ldots, K$. For each combination of factor values $(f_1, \ldots, f_K)$ an indicator variable I can be defined with sample values

$$I_{ij} = \mathbf{1}_{\{F_{ij}^1 = f_1\}} \cdot \ldots \cdot \mathbf{1}_{\{F_{ij}^K = f_K\}}.$$

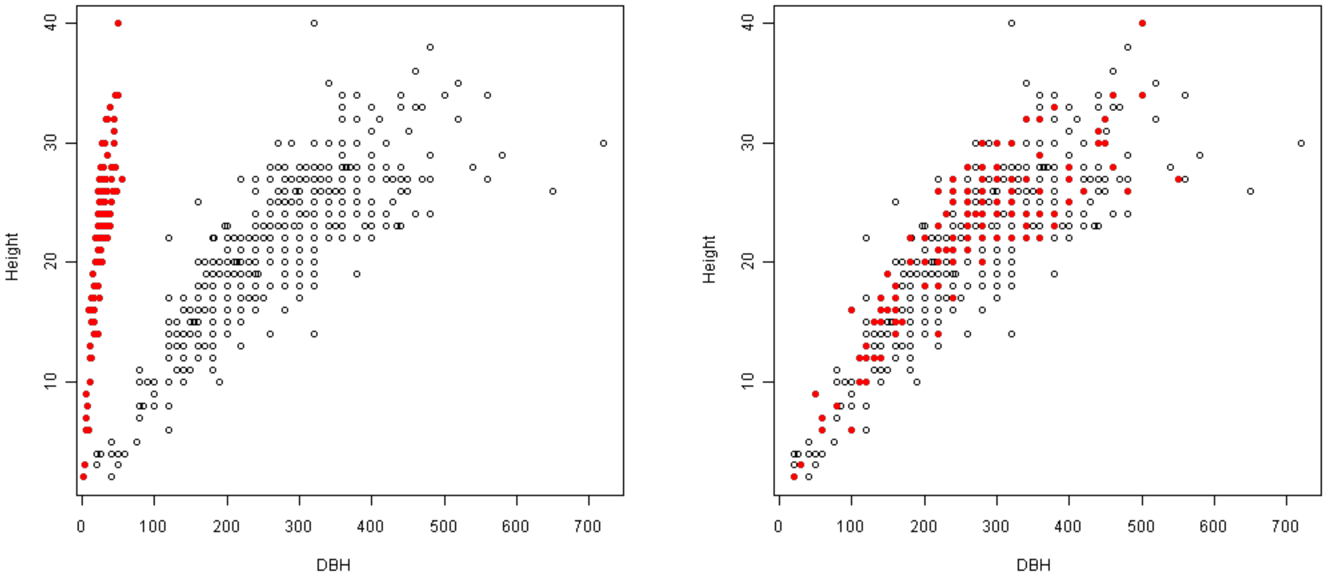
In our inventory, every table is subdivided by one, two or three factors. Below are indicator values for two factors table.

	$F_{ij}^2 = C_1^2$	$F_{ij}^2 = C_2^2$	$\ldots$	$F_{ij}^2 = C_{N_2}^2$
$F_{ij}^1 = C_1^1$	$\mathbf{1}_{\{F_{ij}^1=C_1^1\}}\mathbf{1}_{\{F_{ij}^2=C_1^2\}}$	$\mathbf{1}_{\{F_{ij}^1=C_1^1\}}\mathbf{1}_{\{F_{ij}^2=C_2^2\}}$	$\ldots$	$\mathbf{1}_{\{F_{ij}^1=C_1^1\}}\mathbf{1}_{\{F_{ij}^2=C_{N_2}^2\}}$
$F_{ij}^1 = C_2^1$	$\mathbf{1}_{\{F_{ij}^1=C_2^1\}}\mathbf{1}_{\{F_{ij}^2=C_1^2\}}$	$\mathbf{1}_{\{F_{ij}^1=C_2^1\}}\mathbf{1}_{\{F_{ij}^2=C_2^2\}}$	$\ldots$	$\mathbf{1}_{\{F_{ij}^1=C_2^1\}}\mathbf{1}_{\{F_{ij}^2=C_{N_2}^2\}}$
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
$F_{ij}^1 = C_{N_1}^1$	$\mathbf{1}_{\{F_{ij}^1=C_{N_1}^1\}}\mathbf{1}_{\{F_{ij}^2=C_1^2\}}$	$\mathbf{1}_{\{F_{ij}^1=C_{N_1}^1\}}\mathbf{1}_{\{F_{ij}^2=C_2^2\}}$	$\ldots$	$\mathbf{1}_{\{F_{ij}^1=C_{N_1}^1\}}\mathbf{1}_{\{F_{ij}^2=C_{N_2}^2\}}$

For example, such factors could be $F^1 =$ "$MainSpecies$" and $F^2 =$ "$ForestType$".

6 Execution flow

Variables and characteristics, collected in the forest undergo several transformation to get output tables. There steps are described below.



(a) Input values

(b) Corrected values

Figure 6: DBH measurements correction

6.1 Tables loading

Single inventory dataset is stored in a separate folder, each table is represented by a csv (comma separated values) file. List of currently used tables with variables and characteristics of interest is provided in the file inputTables.csv.

As several yearly inventories are consecutively conducted for the same territory, we can join their records to get improved accuracy of population estimates. Accordingly to the accepted time-space design the estimations for five years period present results for a single NFI circle.

6.2 Data processing

Data processing actions taken are briefly described below.

1.**checkData.r**. To get information about data quality several checks are performed. For example, characteristic values should be inside specific ranges, categorical values should be from predefined set, important characteristics should not be missed.

2.**correctData.r**. Here some common mistakes are fixed, some dummy values are added. Common mistake is measurement of DBH in centimeters instead of millimeters. At the picture 6 filled dots represent erronous measurements, defined by relation between trees DBH and Height. After multiplication of DBH by 10 they become correct.

3.**moveData.r**. Measured values, that are mentioned in one table, are being moved to another table.

4.**prepareIndicators.r**. Indicators for domains selection are computed.

5.**prepareData.r**. Some extra values are compute, including growth level and height estimates.

6.**prepareVolume.r**. Tree volume is modelled using DBH, height and growth level. After that tree volumes are aggregated to the subplots. During that process volumes are compensated by M_s factor of small plots. Thus volume variables for subplots are named as volume_count.

7.**prepareGroups.r**. Categorical groups are computed which will be used as factors in tables.

6.3 Tables computing

After all data has been prepared, target tables can be computed.

Main details of inventory tables are specified in the table 3. A column **Table** lists the numbers of result tables. Here a sign /3 means each third number of the range. **Parameter** is the target value to be estimated. **Object** is object of data collection. The columns **X**, **M** and **Y** are the values to be calculated or estimated. In case, when **Y** is present, ratio **XM/Y** is calculated.

Table	Parameter	Object	X	M	Y
1.(01 - 19), 6.(01 - 03)	Total area	subplot	area	-	-
2.(01 - 17)	Growing stock	subplot	volume_count	-	-
3.(01-16/3, 13.1)	Growing stock	tree	volume	count	-
3.(02-17/3, 14.1), 6.(04, 05)	Number of trees	tree	count	-	-
3.(03-18/3)	Average volume	tree	volume	count	count
5.(01, 03, 10)	Average stock per ha	subplot	volume_count	-	area
5.02	Average age of stand	subplot	stand age	area	area
5.04	Average diameter of stand	subplot	stand diameter	area	area
5.05	Average diameter of trees	tree	diameter	count	count
5.06	Average height of stand	subplot	stand height	area	area
5.07	Average height of trees	tree	height	count	count
5.08	Average site index	subplot	site index	area	area
5.09	Average density	subplot	stand density	area	area
7.01	Total dry volume of stand	subplot	dead volume_count	-	-
7.02	Average dry volume of stand	subplot	dead volume_count	-	area
7.03	Total volume of damaged trees	tree	volume	count	-
7.04	Total volume of dry trees	tree	volume	count	-
7.05	Total volume of deadwood	tree	volume	count	-
8.(01, 02)	Number of regeneration trees	tree	count	-	-

Table 3: Main details of inventory tables

Precise definition of target tables defined in the file outputTables.csv. Meaning of its columns is described in the table 4. To compose a new table, one should add a new line to the file outputTables.csv choosing necessary parameters.

6.4 Examples of area estimations

Let the table 1.01 "Area of forest and non-forest lands" presents a distribution of total area of region in the different categories of lands. In such case X_{ij}, $j = 1, \ldots, N_i$ are areas of sub plots in tract i, and L_{ij} is land category defined for sub plot j in tract i. To estimate total area of category k we define conditional indicator $I_{ij} = \mathbf{1}_{\{L_{ij}=k\}}$. Then formulae (7 - 10) will provide estimates of total area and coefficient of variation with type k.

Computation for other area tables will differ only in indicator definition. For example, only forested land area type is considered in table 1.04 "Area of stands of dominant tree species by users" (i.e. code of Land_Categories_Group = 11 and relative indicator $I_{ij} = \mathbf{1}_{\{L_{ij}=11\}}$. The table has two factors: Main_-Species and Name_of_User. Thus define their values by MS_{ij} and NU_{ij} respectively. Then, estimates for Main_Species m and Name_of_User n can be obtained, using indicators $I_{ij} = \mathbf{1}_{\{L_{ij}=11\}}\mathbf{1}_{\{MS_{ij}=m\}}\mathbf{1}_{\{NU_{ij}=n\}}$ and formulae (7 - 10).

Name	Meaning
number	Table number
main_title	Table title
dimentions	Name of characteristic dimention
data_frame	Field table to be used
condition	Name of indicator variable to specify specific domain
factor1	First factor to specify subdomains
factor2	Second factor to specify subdomains
factor3	Third factor to specify subdomains
table_mode	Table mode, one of the following: - none : compute total or ratio (with two characteristics) values - row : compute percentages for each row - cv : compute coefficients of variation
characteristic1	Name of characteristic
characteristic2	Name of characteristic for denominator
multiplier	Additional constant multiplier
precision	The finest precision to be used for rounding

Table 4: Inventory tables parameters

6.5 Multiple years

Each fifth year inventory is conducted at the same plots. Thus, we can estimate forest dynamics for this period $t_2 - t_1 = 5$ years. Let V_1 and V_2 are the tree volumes at these inventories. Then we can compute such characteristics:

1. **Growth**. For each live tree that tallied at time t_1 and survive until time t_2, we compute annual growth as

$$Growth\_volume = (V_2 - V_1)/(t_2 - t_1)$$

For conservative estimation we neglect the growth of the ingrowth trees that grow across the minimun DBH threshold between time t_1 and time t_2

$$Growth\_volume_I = 0$$

as well as the growth on the mortality trees that died

$$Growth\_volume_M = 0$$

or were cutted between time t_1 and time t_2

$$Growth\_volume_C = 0$$

2. **Cutting volume**. Tree's volume is counted as cutted volume between inventories, if it's diameter in the second inventory became equal to zero. For such trees annual cutting volume is

$$Cutting\_volume = V_1/(t_2 - t_1)$$

3. **Mortality volume**. Tree's volume is counted as mortality volume between inventories, if at the first inventory it was alive, but in the second inventory it was defined as dead. For such trees per year mortality volume is

$$Mortality\_volume = V_2/(t_2 - t_1)$$

By analogy with single year inventory tables main table details are specified in the table 5.

Table	Parameter	Object	X	M
4.01	Per year growth	subplot	Growth_volume_count	-
4.02	Per year growth	tree	Growth_volume	count
4.03	Per year cutting volume	subplot	Cutting_volume_count	-
4.04	Per year cutting volume	tree	Cutting_volume	count
4.05	Per year mortality volume	subplot	Mortality_volume_count	-
4.06	Per year mortality volume	tree	Mortality_volume	count

Table 5: Main details of growth inventory tables

7 Conclusion

Future work:

- Correct missing and incorrect values in the forest. Develop special application for this.
- Improve height and volume models using more complicated formulae, ex. https://academic.oup.com/forestry/article/90/3/359/2605859.
- Fix missing values in factors ('mis' in tables).
- Improve processing of missing values

References

[1] Sampling Techniques, 3rd edition, 1977, W. G. Cochran, Wiley, New York, chapter 6.

[2] Optimal Sample Design for Extensive Forest Inventories, Andrian Lanz, Swiss Federal Institute of Technology, Zurich, 20000.

[3] Volume Assortments Tables for Standing Trees. K. Nikitin, Ministry of Forestry of Ukrainian SSR, Kiev, 631 pages, 1983.