

Kyrgyz National Forest Inventory

Data analysis software design for the Leskhoz management inventory (LMI)

Version 1.2

B. Traub
division Landscape Inventory
WSL

30 November, 2000

1. INTRODUCTION.....	2
2. MODULE 3 DATABASE AND DERIVED ATTRIBUTES	2
2.1. DERIVED ATTRIBUTES FROM DATA ON PLOT LEVEL.....	2
2.2. DERIVED ATTRIBUTES FROM DATA ON TREE LEVEL	4
2.2.1. Relating single tree values to unit area / weighting factors.....	4
2.2.2. Weighting factors for sample plots at the forest edge.....	4
3. MODULE 4 DATA ANALYSIS	5
3.1. METHODS.....	6
3.1.1. Simple random sampling	6
3.1.2. Leskhoz Inventories: Stratified Random Sampling.....	7
3.1.3. Total values.....	9
3.1.4. Hectare related values	10
3.1.5. Combining regional estimates (forest ranges) to Leskhoz estimates.....	11
3.2. DATA PROCESSING / DEFINITION OF ANALYSIS / PARAMETERISATION	12
3.2.1. Analysis Unit.....	13
3.2.2. Reference unit	13
3.2.3. Assessment Units.....	13
3.2.4. Control of data analysis and calculation of estimators	16
3.3. DATA PROCESSING / CONTROL OF ANALYSIS / CALCULATION OF ESTIMATORS	17
3.3.1. Stratification.....	17
3.3.2. Data query / preparing a specific data set according to the parameter settings.....	18
3.3.3. Calculation of terrestrial sample size	19
3.3.4. Estimator calculation.....	21
3.3.4.1. loop over assessment units.....	22
3.3.4.2. summarizing on Leskhoz level.....	25
4. REFERENCES.....	26
5. APPENDIX 1 FLOW CHARTS.....	27
6. APPENDIX 2.....	29
6.1. SAMPLE SIZE	29
6.2. DEFINITIONS AND NOTATION	29
6.3. EXPORT FILE.....	31
6.4. TEMPORARY TABLES.....	32
7. APPENDIX 3 DATA BASE TABLES.....	33

1. Introduction

The concept of the Kyrgyz National Forest Inventory is based on two related inventory systems: the national forest inventory (NFI) and the Leskhoz management inventory (LMI). The NFI plots are distributed over the whole country and match to the 500m x 500m inventory grid of the Leskhoz. The definitions and measurement rules of the attributes to be assessed are harmonised. Data from both inventories are stored in a common database and will be processed and analysed by an integrated software tool.

This report is based on Köhl and Scheuber, 1999. (algorithms for estimator and variance calculation / description of the statistical methods) and also on Traub 1999 (data analysis concept). The definitions and measurement rules of the field survey are given in Scheuber 2000 (LMI manual).

In general the data analysis software tool provides results for geographical subunits of the Leskhoz, like forest ranges (Lesnichestvo). A further subdivision in assessment units allows a differentiation of the results according to thematic categories, like ownership category or forest type.

As it was agreed in former missions, the attributes assessed for the LMI are stored as raw, non-aggregated data in a relational database management system (DBMS). Data aggregation steps are conducted in module 3 and 4 (see Fig. 2.1) The analysis software (module 4) controls the access to the database, the data flow and data processing steps, i.e. the statistical procedures or estimator calculation etc..

The database query and the calculation of the estimators consume considerable DBMS resources. Thus the parameterisation of the data analysis, the database query and data processing steps as well as the presentation of results are carried out separately. Whereas the data query is stored only temporarily, the data analysis parameters and the results of estimator calculation are stored permanently. This procedure avoids data redundancy and allows to carry out and to reproduce an analysis topic several times.

Subsequently the modules 3 and 4 of Fig. 2.1 are described and explained. Module 1 and 2 were already prepared by the Kyrgyz partner (GIS Lab). The proposed design does not depend on certain hard- and software resources and tools. The data query step description however follows the structured query language (SQL / ORACLE sqlplus). So, the implementation of the procedures is to be realised by the Kyrgyz partners according to their resources and current situation.

2. Module 3 Database and derived attributes

For the analysis of inventory data and for the calculation of estimators several so called 'derived attributes' have to be derived from raw data. They could be stored permanently in the inventory database or optionally calculated during runtime. The latter method is recommended if the derived attribute combines different type of data (plot/subplot data, data dependent from inventory type or occasion, stratification variables)

2.1. derived attributes from data on plot level

In general this group comprises classified metric data and reassigned categorical data like e.g. a binary grouping of all ownership categories in the classes 'private' and 'public'.

In particular plot level data concern:

- data from GIS, georeferenced point data
- Site property data
- Slope data ($\text{slope} = (\text{abs}(\text{slope1}) + \text{abs}(\text{slope2})) / 2$)
- Reduced plot size for plots intersecting the forest margin

- Degree of mixture (based on the proportion of basal area)
- land use class (**plot.luc**) respectively forest type/forest definition derived from the attribute **plot.landuse**

	plot.luc	definition of forest in a strict sense plot.landuse	definition of forest in a wider sense plot.landuse
non-forest	1	5	5
forest	2	1,4	1,2,3,4
other wooded land	3	3	
purposeful plantation	4	2	

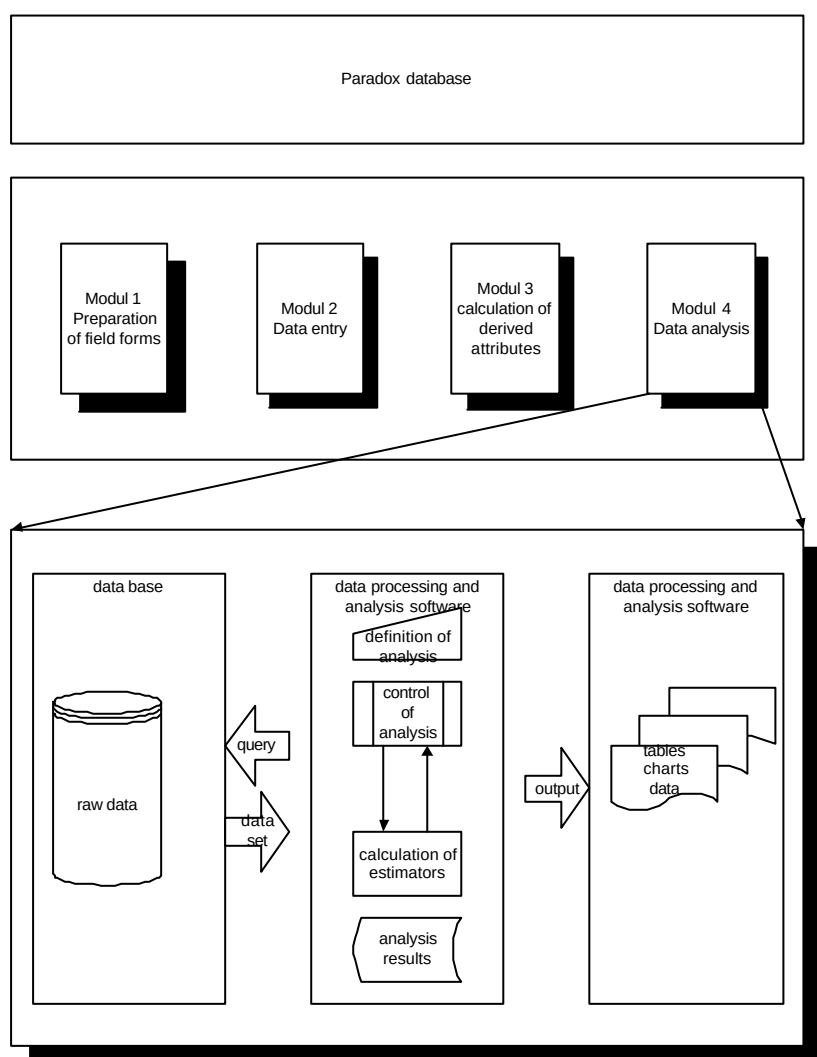


Fig. 2.1:Kyrgyz National Forest Inventory (NFI) and Leskhoz Management Inventory (LMI) data analysis concept.

2.2. derived attributes from data on tree level

- Basal area (transformation of DBH)
- Single tree volume (to be developed by co-operation of TU Dresden and Forest Institute Bishkek)
- Assortment volume (see single tree volume)
- Increment (By applying the increment tariffs (increment as a function of DBH) (Mission report: Köhl, Scheuber, Traub 13.9.99, Appendix B).
- Expansion factor w_{ij}

2.2.1. Relating single tree values to unit area / weighting factors

Since the sample plots are the random sample units of the terrestrial assessment, the single tree attributes must be aggregated/combined to sample plot values.

The values of the single tree attributes are assessed on circular sample plots and thus must be related to a unit area (1 hectare) in order to consider the selection probability of a tree. For this purpose individual weights w_{ij} are assigned to each single tree on plot i . The weighted attributes Y_{ij} are summarised to plot values

Herewith also the correction due to the selection probability for sample plots at the forest edge is done.

$$w_{ij} = A / a_{ij} = 1ha / a_{ij} \quad eq. 2.1$$

where:

A = reference area (Typically, the unit area for forest inventories is 1 ha (10'000 m²)).

a_{ij} = area of sample plot i on which tree j is located.

Since concentric sample plots are assessed, different weighting factors result for the 200 m² and 500 m² - plots. For sample plots, which do not extend beyond the forest edge, the weights are constant.

$a_{ij}=0.2ha$ for trees with $12 < DBH < 36$ $\rightarrow w_{ij} = 50$

$a_{ij}=0.5 ha$ for trees with $DBH > 36 cm$ $\rightarrow w_{ij} = 20$

2.2.2. Weighting factors for sample plots at the forest edge

Sample plots intersecting the forest edge are assessed if the plot centre is located inside the forest area.

For all trees on these plots ,individual' expansion factors are to be applied, which are specified by a surface a_i , which contains all localisation's of sample plot centres (Y, X). The surface a_i is called representation area in the following. In case of concentric circular sample plots, different weighting factors exist depending on the defined minimum diameter of the plots.

If the distance of a tree from the forest edge is smaller than the radius of the sample plot, the tree is called edge tree and the representation area a_i must be reduced by the part, which lies outside of the forest area. As the probability of selection (p_i) of a tree is specified through

$$p_i = a_i / A \quad eq. 2.2$$

where

$A = 10'000 [m^2]$ and

a_i = size of plot i [m^2]

In Fig. 2.2, the steps of data aggregation and estimator calculation is given schematically. Based on **plot-related** (e.g. stand development stage, ownership status, etc.) and **single tree-related attributes** (e.g. timber volume, stem damages) **total Y_j values** (e.g. total timber volume per tree species) are calculated. In a subsequent step the forest area is estimated, which likewise leads to **total X_j values**. **Ratio estimation** (total Y_j / total X_j) leads to area related estimators such as volume per ha or to percentages (e.g. young forest area in proportion to the total forest area). This ratio estimators must not be mixed up with the unit-area related ‘per ha’ values for single tree attributes.

The estimates are at first calculated on the regional level i.e. separately by forest ranges. In a following step the regional values –which are treated as independent samples- are combined to values for the whole Leskhoz.

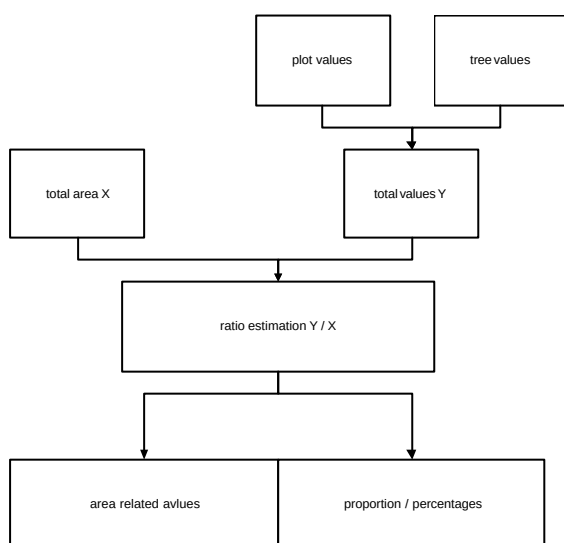


Fig. 2.2: hierarchy of data aggregation and estimator calculation

3. Module 4 Data analysis

The LMI-data analysis software (Module 4 in Fig. 2.1) is structured in three main parts:

1. Database access facilities and interfaces
2. Data processing and analysis software: definition/parameterisation, data-flow control, estimator calculation
3. Data processing and analysis software: reporting software

In the following the statistical procedure to calculate the Y and X variables (target variables and reference unit) is presented. The algorithms of (stratified) random sampling theory to calculate totals, ratio estimators and sampling errors are applied to the systematic sampling approach of the LMI. This leads to an overestimation and thus to a conservative estimation of the sampling error.

3.1. Methods

3.1.1. Simple random sampling

In case of simple random sampling (rs) each element of the sampled population has an equal probability to be selected. The average value of y is computed as follows:

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n} \quad \text{eq. 3.1}$$

the variance s^2 is calculated as:

$$s^2 = \frac{\sum (y_i - \bar{y})^2}{n-1} = \frac{\sum y_i^2 - (\sum y_i)^2}{n(n-1)} \quad \text{eq. 3.2}$$

(the second term is more convenient for computation, see also chapter 6.2)

the standard error of the mean $s_{\bar{x}}$ finally results in:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} \quad \text{eq. 3.3}$$

with

n = sample size (number of samples)

$\bar{y}$ = average value

y_i = value of element i in trial plot j

s^2 = variance

$s_{\bar{x}}$ = standard error of the mean

If proportions are to be calculated the following formulas hold:

$$\bar{y} = \frac{\sum y_i}{n} = \frac{a}{n} = p \quad \text{eq. 3.4}$$

with

$$y_i = \begin{cases} 1, & \text{if the element is in the considered class} \\ 0, & \text{in all other cases} \end{cases}$$

a = number of elements in the considered class.

This applies to target variables in a categorical data format where the area portion of a certain categories is estimated (e.g. portion of forest area assigned to a certain development stage)

The variance of this proportion estimate is equal to eq. 3.2, however can be expressed also by

$$s^2 = \frac{n}{n-1} pq \quad \text{eq. 3.5}$$

with $q = 1-p$.

The standard error is equivalent to eq. 3.3.

3.1.2. Leskhoz Inventories: Stratified Random Sampling

For the Leskhoz inventories stratified random sampling is applied. Stratification is efficient if the population is divided in homogeneous parts (effect: low variance within the strata and high variance between the strata). Efficient stratification strategies thus lead to a reduced sample size for a given precision that is to be achieved.

In the LMI the strata are built according to the attribute **stand.stgroup** (stand type) and **plot.luc** (landuse class). The stratum assignment results in a derived attribute calculated by

$$\mathbf{stratum} = 10 * \mathbf{luc} + \mathbf{stgroup}$$

For certain plots assigned to forest no stand information might be available (due to storm damage, clearcut etc.). Thus the attribute **stratum** is stored in table **plot**. Alternatively the variable **stratum** can be built during analysis runtime.

LMI results are calculated separately for regions i.e. forest ranges (**general.unit2**) and as well for the entire area of a Leskhoz. The regions may be divided in assessment units (cells of result tables).

The number of samples in a stratum is proportional to the size of the stratum, because the samples are distributed systematically within a fixed 500 x 500 meter grid. The number of samples in the strata N_h , and the total sample size N can be derived directly from the inventory database. In this context it is important to count all plots without any restrictions concerning the Y variable.

The average value of a variable y per stratum ($\bar{y}_{st}$) is calculated as:

$$\bar{y}_{st} = \frac{\sum_{h=1}^L n_h \bar{y}_h}{n} \quad eq. 3.6$$

since $\frac{n_h}{n} \approx \frac{A_h}{A}$ for area proportional sample distribution in case of systematic sampling, $\bar{y}_{st}$ can be calculated by:

$$\bar{y}_{st} = \frac{\sum_{h=1}^L A_h \bar{y}_h}{\sum_{h=1}^L A_h} = \frac{\sum_{h=1}^L A_h \bar{y}_h}{A} \quad eq. 3.7$$

where

$$\bar{y}_h = \frac{\sum_{i=1}^{n_h} y_{ih}}{n_h} \quad (\text{mean in stratum } h)$$

A_h = area of stratum h (could be estimated by the proportion of terrestrial sample size in stratum h)

A = total area (known)

n_h = sample size in stratum h

n = total sample size ($n = \sum_{h=1}^L n_h$)

The variance of the mean $v(\bar{y}_{st})$ for large N is computed by:

$$v(\bar{y}_{st}) = \sum_{h=1}^L \frac{w_h^2 s_h^2}{n_h} \quad eq. 3.8$$

with:

$$w_h = \frac{n_h}{n} \approx \frac{A_h}{A} \text{ (for area proportional sample plot distribution)} \quad eq. 3.9$$

and:

$$s_h^2 = \frac{\sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)^2}{n_h - 1} = \frac{\sum_{i=1}^{n_h} y_{hi}^2 - \left(\sum_{i=1}^{n_h} y_{hi} \right)^2 / n_h}{n_h - 1} \quad eq. 3.10$$

Area portions: (applied to categorical target variables, e.g. area portion of a certain degree of mixture)

$$p_{st} = \sum_{h=1}^L \frac{n_h p_h}{n} = \sum_{h=1}^L \frac{A_h p_h}{A} \quad eq. 3.11$$

with:

$$p_h = \frac{a_h}{n_h} = \text{portion in stratum } h$$

a_h = number of elements in the considered class of target variable Y in stratum h

This means that in case of a categorical target variable the following subroutine has to be evaluated (see chapter 3.3.4 and Fig. 3.3)

$$count \text{ var} = \begin{cases} 1 & \text{if } kateg = i \\ 0 & \text{if } kateg \neq i \end{cases}$$

Where the variance $v(p_{st})$ of the estimated portion is calculated as:

$$v(p_{st}) = \sum_{h=1}^L w_h^2 \frac{p_h q_h}{n_h - 1} \approx \frac{1}{n} \sum_{h=1}^L w_h p_h q_h \quad eq. 3.12$$

proof: see chapter. 6.2

It is important to differ between the term *stratum* in the sense of stratified sampling and the term *stratum* in the sense of an assessment unit. Assessment units are defined as the smallest unit for estimator

calculation; they can cover 1 or more strata! If strata and subunit definition is equivalent, both units are congruent, i.e. each subunit covers exactly one stratum ($L=1$ for each assessment unit).

The total forest perimeter in the LMI is available from GIS. In other cases it is recommended to estimate the forest area by dot grid sampling. The target values of area estimation could be:

- 1) the portion of forest area,
- 2) the absolute forest area.

The absolute forest area depends on the number of points on the category forest and the spacing of the dot grid. The area portion is derived from the relationship of dots located in forest area to the total number of dots.

The forest area portion p results from:

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n} = \frac{n_w}{n} = p \quad \text{eq. 3.13}$$

with

n = number of random points and

n_w = number of points that represent forest.

p is an estimation of the true portion π in the population. The estimated portion of non-forest area is named q and follows after:

$$q = 1 - p \quad \text{eq. 3.14}$$

The variance of p , $v(p)$, is:

$$v(p) = s(p)^2 \cong \frac{pq}{n-1} \quad \text{eq. 3.15}$$

The standard error $s(p)$ and relative standard error SE% are calculated as:

$$s(p) = \sqrt{v(p)} \quad \text{eq. 3.16}$$

$$SE \% = \frac{s(p)}{p} 100 \quad \text{eq. 3.17}$$

Note that the variance of the estimated portion depends only on p and the sample size n . With fixed n the variance is greatest if $p = q = 0.5$ and diminishes asymptotically for $p = 0$ or $p = 1$.

3.1.3. Total values

The total value of a certain attribute k (e.g. timber volume) is obtained by the multiplication of the average value with the total area of the population.

$$\hat{Y}_k = A_T \hat{y}_k \quad \text{eq. 3.18}$$

where

A_T = area of the region in hectares.

In this context it is crucial, how the total area of the population is defined. On the assumption of a priori determined sample sizes the value zero is assigned to attributes, which are not observed on a sample plot. On non-forest sample plots thus all attributes adopt the value zero. Thus the total area of the population A_T contains forest areas as well as non-forest areas. This assumption is insignificant for the derivation of total values, as the following example shows (Tab. 3.1).

Tab. 3.1 Sample size and attribute values

	Area	Number of Samples	Attribute Value (per hectare)	Total Value (area * attribute value)
Forest	5000 ha	50	100 m ³ (/ha)	500000 m ³
Non-Forest	5000 ha	50	0 m ³ (/ha)	0 m ³
Total	10000 ha	100	50 m ³ (/ha)	500000 m ³

The variance of the total value depends again on the design. The following estimator is valid for simple random sampling and leads to conservative (to high) estimates in case of systematic sampling.

$$v(\hat{Y}_k) = A_T^2 v(\hat{y}_k) \quad \text{eq. 3.19}$$

It is assumed that A_T can be derived from GIS data sources or official statistics etc.. Therefore no statistical estimation error component for A_T needs to be involved for the computing the total values.

3.1.4. Hectare related values

In forestry practice attribute values are commonly related to one hectare (e.g. volume in m³/ha, basal area [m²/ha] etc.). These values are derived by ratio estimators (quotient of the total value of the target attribute, Y_k , and the area of the reference unit, A_k ; (to be derived from GIS or estimated from the sample)

$$\hat{R}_k = \frac{\hat{Y}_k}{\hat{A}_k} = \frac{\hat{\bar{y}}_k}{\hat{\bar{A}}_k} = \frac{\sum \frac{\hat{y}_{ik}}{n}}{\sum \frac{\hat{A}_{ik}}{n}} \quad \text{eq. 3.20}$$

$\hat{R}_k$ is the of ratio of means for attribute k , which results for large sample size in unbiased estimates.

The variance of $\hat{R}_k$ is calculated by

$$v(\hat{R}_k) = \frac{v(\hat{Y}_k) + R_k^2 v(\hat{A}_k) - 2\hat{R}_k \text{cov}(\hat{Y}_k, \hat{A}_k)}{n\hat{A}_k^2} \quad \text{eq. 3.21}$$

with

$$\text{cov}(\hat{Y}_k, \hat{A}_k) = A_T^2 \frac{\sum_{i=1}^n (y_{ik} - \hat{y})(A_{ik} - \hat{A}_{ik})}{n(n-1)} \quad \text{eq. 3.22}$$

The co-variance term accounts for the dependence of the target variable (nominator) and the area estimation (denominator) (Cochran, 1977).

The variances $v(\hat{Y}_k)$ and $v(\hat{A}_k)$ are computed by eq. 3.19.

Attribute values may also be related on reference units such as number of stems or to the total values of the attributes (calculation of portions). In this cases the special attribute $\hat{A}_k$ is replaced by $\hat{X}_k$ as it was generally applied in chapter 3.3.

3.1.5. Combining regional estimates (forest ranges) to Leskhoz estimates

The value of an attribute on the Leskhoz level is formed by summing up the total values of the individual regions.

$$\hat{Y}_{Nk} = \sum_{l=1}^s \hat{Y}_{lk} \quad \text{eq. 3.23}$$

$$\hat{X}_{Nk} = \sum_{l=1}^s \hat{X}_{lk} \quad \text{eq. 3.24}$$

where $\hat{Y}_{lk}$ and $\hat{X}_{lk}$ are regional values for attribute k . $\hat{Y}_{Nk}$ and $\hat{X}_{Nk}$ are the total values for the Leskhoz (s = number of forest ranges).

Since this is the sum of s independently assessed variables, the variance of the total values is calculated by

$$v(\hat{Y}_{Nk}) = \sum_{l=1}^s v(\hat{Y}_{lk}) = \sum_{l=1}^s A_{lk}^2 v(\hat{Y}_{lk}) \quad \text{eq. 3.25}$$

$$v(\hat{X}_{Nk}) = \sum_{l=1}^s v(\hat{X}_{lk}) = \sum_{l=1}^s A_{lk}^2 v(\hat{X}_{lk}) \quad \text{eq. 3.26}$$

On the Leskhoz level hectare-related estimates are calculated with an identical approach as used for regions.

$$\hat{R}_{Nk} = \frac{\hat{Y}_{Nk}}{\hat{A}_{Nk}} \quad \text{eq. 3.27}$$

$$v(\hat{R}_{Nk}) = \frac{v(\hat{Y}_{Nk}) + \hat{R}_{Nk}^2 v(\hat{X}_{Nk}) - 2 \hat{R}_{Nk} s_{YXNk} / n}{\hat{X}_{Nk}^2} \quad \text{eq. 3.28}$$

with

$$s_{YXnk} = \sum_{l=1}^s A_K^2 s_{YXlk}, n = \text{total number of plots in the Leskhoz} \quad \text{eq. 3.29}$$

3.2. Data processing / Definition of analysis / Parameterisation

The organisation of a correct parameterisation is ensured in two different ways. (1) the items and selection fields are adapted to the parameters that have already been set. (2) tests are conducted to examine the validity of the parameter set-up. Thus incorrect analyses can be mostly avoided.

The contents of the parameterisation window are given in Tab. 3.2

Tab. 3.2.: main set of parameters for the LMI data analysis

<i>topic</i>	link to a (internal) table topics where the whole set of data analysis definitions is stored. This table allows to reproduce/modify a certain analysis.
<i>inventory perimeter (ip)</i>	determines the perimeter to be analysed, i.e. the Leskhoz in case of LMI analysis or total area of Kyrgyzstan for NFI analysis. (general.unit1 , LMI manual 2.5).
<i>inventory cycle (ic)</i> <i>inventory occasion</i>	It is possible to analyse a state or change assessments or to choose different type of inventories (n.b.: at the moment only LMI state assessment is applied)
<i>analysis unit (aunit)</i>	The data analysis usually refers to the accessible forest without shrub forest. Condition: plot.landuse =2 and plot.access = yes (LMI manual 1.16 and 1.15). The analysis unit defines the sampled population / i.e. determines the number of terrestrial sample plots to be included in the analysis (c.f. ch. 3.2.1).
<i>inventory unit (invunit)</i>	determines the regional subdivision of the <i>inventory perimeter</i> (here division of Leskhoz in forest ranges / Lesnichestvo general.unit2 (LMI manual 2.6)). The inventory units are geographically well defined regions with known area. The estimates are calculated separately for each inventory unit.
<i>condition</i>	condition to restrict the analysis data set. For certain topics (e.g. timber volume, increments, and utilisation) it is, at times, necessary to restrict the data set to be analysed. Typically certain types of trees or plots need to be excluded from the analysis. Important: the restriction applies only to the target variable, it must not reduce the sample size which is determined by the choice of the <i>assessment units</i>
<i>target variable</i>	attribute to be analysed. The analysis is adjusted depending on the type (continuous or categorical) and the level (sample plot or subunit) of a target variable. The classification of metric target variable or the reassignment of categorical variables can be done optional.
<i>reference unit (ru)</i>	the <i>reference unit</i> (ha, no of stems, volume, percent, c.f. ch. 3.2.2) can be selected independently of the target variable. The differentiation of the reference unit (calculated separately for each table cell or uniformly for columns, rows, or the entire table) however depends on several conditions (c.f. ch. 3.2.2)
<i>NVL</i>	determines the treatment of missing values (null values). By default they are replaced during the database query by the value zero (0).
<i>assessment unit (au)</i>	smallest unit of estimator calculation (resolution of results/table cells, c.f. Fig. 3.1). The categorising of the assessment unit variables can be processed

	interactively or by lookup-tables.
--	------------------------------------

3.2.1. Analysis Unit

The selection of the population to be analysed determines the sample size and the target variables available for the analysis as well. In Tab. 3.3 the different categories of analysis units are given

Tab. 3.3: Definition of the analysis unit.

unit	Description	Condition
<i>total</i>	total area, useful for analysis of non-forest specific data. sample size: total no of plots on the 500m grid.	none
<i>accessible forest</i>	major unit to analyse LMI data: accessible forest plots without shrub forest. sample size: accessible forest plots	plot.luc = 2 and plot.access = yes
<i>shrubs forest</i>	analysis of shrub forest. sample size: shrub forest plots	plot.luc = 3
<i>young growth</i>	young growth analysis. sample size: no of accessible forest plots where young forest was assessed	plot.luc = 2 and plot.access = yes and yfnature.type is not null
<i>purposeful plantation</i>	Purposeful plantation survey. sample size: no of plots where purposeful plantation were assessed	plot.luc = 4 and plot.access = yes

3.2.2. Reference unit

The reference unit of the analysis is used for the calculation of ratio estimates (denominator/*X* variable). For LMI data analysis typically the attributes *forest area*, *number of stems* and *proportions* are applied. Furthermore, the design of the reference-unit-table (degree of differentiation) can be selected. The allowed options depend (1) on the level of the *target attribute*, (2) the *assessment units* and (3) the level of the *reference unit* chosen (c.f. Fig. 3.1).

Options:

Reference to a cell	an individual reference unit is derived for each cell	(1)
Reference to a column	one common reference unit is derived for each column	(2)
Reference to a row	one common reference unit is derived for each row	(3)
Reference to a table	one common reference unit is derived for the whole table	(4)

3.2.3. Assessment Units

The highest 'resolution' of data analysis are assessment units. They are built according to the categories of one or two attribute(s). The number of rows and columns of the result table and thus the sample size for the assessment units is determined by the number of categories of these attributes.

example:

The analysis of the target variable *timber volume* (level: subplot) by the *assessment units*:

(1) *ownership* (level: plot), and

(2) *tree species* (level: subplot),

is to be carried out

Further, the results should be related to the *reference units*:

(a) *forest area* (level: plot) and

(b) *number of stems* (level: subplot):

to (a), reference unit = forest area

The timber volume per hectare can either be calculated for each class of the variable *ownership* or for the entire *forest area*. The result 'Timber volume per hectare of forest area with occurrence of a certain *tree species*' is, however, not very informative and results in non-additive tables. Thus the differentiation of the reference unit *forest area* is allowed for the assessment unit *ownership* but not for the assessment unit *tree species*.

to (b), reference unit = number of stems

In this case all possible types of reference unit tables are allowed. It is also possible to calculate the mean volume per tree according to: 1) *ownership* and 2) *tree species*.

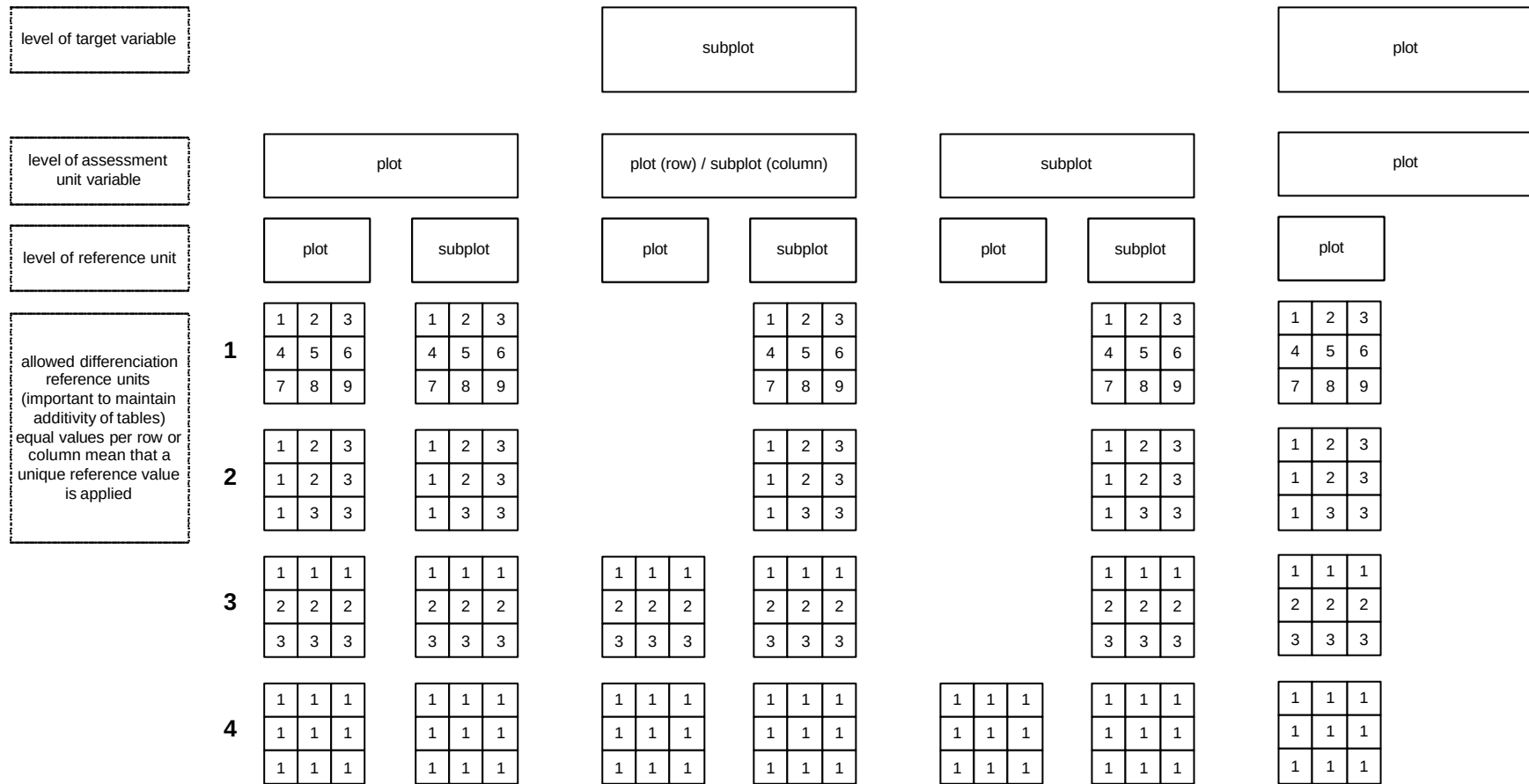


Fig. 3.1: Scheme of possible reference unit tables (c.f. chapter2, figure 3). The same number in the cells of the individual tables mean equal reference unit figures. **1**=cell, **2**=column-, **3**=row-, **4**=table related reference unit.

3.2.4. Control of data analysis and calculation of estimators

The process of the data analysis is presented in a simplified form in Fig. 3.2. For the simple random sampling analysis, the estimates are directly derived for each assessment unit. For the stratified random sampling in a first step the estimators are calculated separate by the strata of a certain assessment unit. In a next step a weighted sum according to the stratum size is built to get a unique stratified estimator per assessment unit.

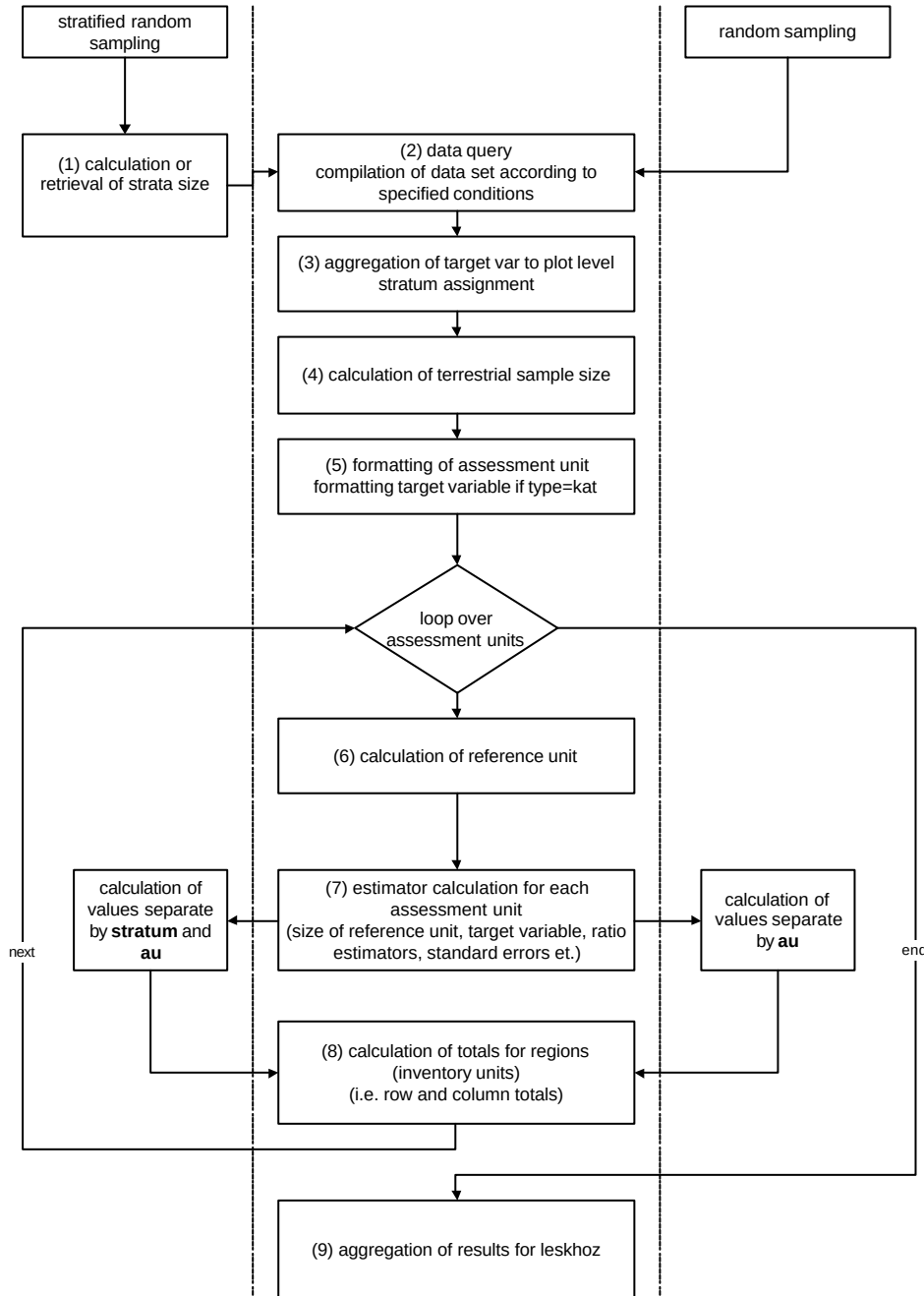


Fig. 3.2: data processing flow. For more detail concerning step5 see Fig. 5.1, for steps 5 – 9 see Fig. 3.3.

3.3. Data processing / control of analysis / calculation of estimators

In this chapter the steps to prepare aggregated data to be presented in result tables are described.

Inventory results are typically provided in tabular form. The columns and rows of the table represent the assessment units. The target variable (e.g. timber volume) determines the thematic contents of the table.

Three types of tables are to be differentiated: **attribute tables**, **reference unit tables** and **ratio estimator tables** (c.f. Köhl, Scheuber, Traub report 9/99, Appendix1, table 4)

The calculation of the reference units is described in chapter 3.2.2

Tab. 3.4: conventions

<i>italics:</i>	in text: fictive table or attribute / variable name
<code>courier 8pt:</code>	SQL- or other programming code, name of tables and variables of the LMI database, temporary tables and variables created during runtime
<code><...>:</code>	facultative
<code>[]:</code>	must have code (applied in some cases for clarification)
<code>/**/:</code>	comments in programming code
<code>%name<param.>:</code>	call of a function (facultative with parameters)

3.3.1. Stratification

For the LMI it is assumed that the relevant information for stratification is available from GIS (stand type coverage). A table or temporal view **strataarea** where the variable **Ah** i.e. the area of stratum *h* in ha for each inventory unit is stored (see eq. 3.7). The key variables are: *invid*, *unit2*, *stratum*.

The variable *stratum* contains the code assigned to each sample plot defined by the components **stand.stgroup** and **plot.luc**; *stratum* = (10*(**plot.luc**+**stand.stgroup**)). This derived variable is not stored permanently but is build during runtime to keep flexibility for the analysis.

Alternatively the stratum area can be estimated from the proportion of the sample size in stratum *h* and the total sample size times the total area of the inventory unit. This is the case if the stratum assignment has to be derived from the inventory data (poststratification)

The following steps are to be carried out in the course of stratification:

(1) calculation of the total area (*A*) of an inventory unit (sum over strata) (see eq. 3.7).

```
create table unitarea as select unit2, sum(ah) a
from strataarea
group by unit2
```

(2) calculation of the stratum weight (*ahn*) which is equivalent to *wh* (see eq. 3.9)

```
create table ahn as select s.unit2, s.stratum, s.ah, a.a, s.ah/u.a ahn
from strataarea s, unitarea u
where s.unit2=u.unit2;
```

Note: *ah/a* sums up to 1 for each inventory unit

table *ahn* (example)

unit2	stratum	ah	a	ahn
1	11	11677	19751	0.5912105716
1	12	14	19751	0.0007088249
1	13	6	19751	0.0003037821
1	14	1	19751	0.0000506303
1	21	18	19751	0.0009113463
1	22	1164	19751	0.0589337249

1	23	4331	19751	0.2192800365
1	24	2403	19751	0.1216647258
1	25	114	19751	0.0057718597
1	32	1	19751	0.0000506303
1	41	22	19751	0.0011138677
sum		19751		1.00
2	11	28548	37645	0.7583477221
2	12	3	37645	0.0000796919
2	13	7	37645	0.0001859477
...	...	...	...	...

(3) general stratum assignment for all plots located in the inventory perimeter (Leskhoz)

table *tmpstrat* (example)

plotid	STRATUM
1	24
2	23
3	24
4	23
5	23
6	23
7	11
8	11
9	11
10	23
11	24
12	23
13	11
14	23
15	23

```
create table tmpstrat as
select g.clnr,
(10*(plot.luc+stand.stgroup)) stratum
from general g, stand s, plot p
where p.invnr = 1
and g.clnr = p.clnr
and g.clnr = s.clnr
and s.invnr=p.invnr
order by g.clnr
```

3.3.2. Data query / preparing a specific data set according to the parameter settings

(1) data query, result: table *tem1*

```
create table tem1 as select
g.clnr, g.x, g.y, g.clnr, g.unit2, target var, au1 var, au2 var <,wij (expansion
factor if target variable is a single tree attribute)>
from general g, plot p, stand s <,further query tables, depending on au's and
target var>
where cl.clnr = cl2.clnr
and g.clnr = p.clnr
and g.clnr = s.clnr,
and 1=p.invnr,
and p.invnr = s.invnr
and <links with further query tables>
and <conditions that define the analysis unit i.e. p.luc and p.access restriction>
and <further conditions respectively restrictions concerning the target var or au
1 / 2 variables>
```

table *tem1* (example target var: single tree volume (type = metric), au1: *altitude* (type = metric), au2: *treetype* (broadleaved /coniferous type = nominal), *x/y* = plot coordinates, *wij* = weight of tree *j* on plot *i* (see eq. 2.1))

X	Y	plotid	treeid	unit2	treetype	altitude	volume	WIJ
686	295	6	281487	1	2	745.4	-1.388	20
686	295	6	281488	1	1	745.4	1.181	50
686	295	6	281489	1	2	745.4	0.685	20
686	295	6	281490	1	2	745.4	0	50
686	295	6	281493	1	2	745.4	3.416	20
686	295	6	281494	1	2	745.4	0.12	50
685	294	22	79290	1	2	844.1	0.39	50
685	294	22	79291	1	1	844.1	0.899	50
685	294	22	79296	1	1	844.1	0.949	50
685	294	22	79297	1	1	844.1	0.677	50
685	294	22	79299	1	2	844.1	0.64	50
685	294	22	79303	1	2	844.1	0.271	50
685	294	22	79304	1	1	844.1	1.348	20
685	294	22	79306	1	2	844.1	0.076	50
685	294	22	79307	1	1	844.1	1.434	20

(2) code tables, result: tables *keytab1/2*

These tables contain the codes of all classification carried out: The codes are queried from the database or lookup-tables. Alternatively an interactive assignment of codes by the user is possible (GUI); if the type of the assessment units chosen is metric in either case interactive definition is required.

table *keytab1* for au1 (example for interactive classification)

altitude	lower	upper	cat1	TEXT
1	0	200	1	1) 1 - 200
2	200	400	2	2) 201 - 400
3	400	600	3	3) 401 - 600
4	600	800	4	4) 601 - 800
5	800	1000	5	5) 801 - 1000
6	1000	1200	6	6) 1001 - 1200
7	1200	1400	7	7) 1201 - 1400
8	1400	1600	8	8) 1401 - 1600
9	1600	1800	9	9) 1601 - 1800
10	1800	2000	10	10) 1801 - 2000
11	2000	2200	11	11) 2001 - 2200
12	2200	2400	12	12) 2201 - 2400

cat = category, text = description of category

table *keytab2* for au2 (example)

treetype	cat2	TEXT
0	1	missing val.
1	2	coniferous
2	3	broadleaved

(3) merging of tables *tem1*, *keytabs*, *tmpstrat* to complete basic table for data analysis, result: table *terr1*.

table *terr1* (example)

X	Y	plotid	treeid	unit2	STRATUM	cat1	cat2	TXT1	TXT2	volume	WIJ
686	295	6	281493	1	23	3	4	2	4) 601 - 800	3.416	20
686	295	6	281492	1	23	3	4	2	4) 601 - 800	0.159	20
686	295	6	281491	1	23	3	4	2	4) 601 - 800	2.181	20
686	295	6	281494	1	23	3	4	2	4) 601 - 800	0.12	50
686	295	6	281487	1	23	3	4	2	4) 601 - 800	-1.388	20
686	295	6	281489	1	23	3	4	2	4) 601 - 800	0.685	20
686	295	6	281488	1	23	2	4	1	4) 601 - 800	1.181	50
686	295	6	281490	1	23	3	4	2	4) 601 - 800	0	50
685	294	22	79309	1	23	3	5	2	5) 801 - 1000	1.323	50
685	294	22	79308	1	23	2	5	1	5) 801 - 1000	0.557	50
685	294	22	79290	1	23	3	5	2	5) 801 - 1000	0.39	50
685	294	22	79303	1	23	3	5	2	5) 801 - 1000	0.271	50
685	294	22	79306	1	23	3	5	2	5) 801 - 1000	0.076	50
685	294	22	79296	1	23	2	5	1	5) 801 - 1000	0.949	50
685	294	22	79307	1	23	2	5	1	5) 801 - 1000	1.434	20

3.3.3. Calculation of terrestrial sample size

In this context it is important to include all sample plots. The restrictions active for the preparation of the analysis data set (c.f. ch. 3.3.2) must not be applied

```
create table nterr
(select stratum, a.invunit, nht, nt, (nht/nt) nhnt
 from (subquery) a,(subquery) z
 where a.invunit = z.invunit
 order by a.invunit, a.stratum);
```

description of subquery alias **a**: (*nht*: no of plots per inventory unit and stratum)

```
(select stratum, g.unit2 invunit, count(*) nht
 from general g, stand s, plot p
 where p.invr = 1
 and g.clnr = p.clnr
```

```

    and g.clnr = s.clnr
    and s.invnr = p.invnr
    and p.luc = 2
    group by g.unit2, stratum) a,

```

description of subquery alias **Z**: (*nt*: no of plots per inventory unit)

```

(select g.unit2 invunit, count(*) nt
 from general g, stand s, plot p
 where p.invnr = 1
    and g.clnr = p.clnr
    and g.clnr = s.clnr
    and s.invnr = p.invnr
 group by g.unit2) z

```

table *nterr* (example)

STRATUM	unit2	NHT	NT	NHNT
21	1	10	1109	0.00901713255184
22	1	195	1109	0.17583408476104
23	1	541	1109	0.487826871055
24	1	326	1109	0.29395852119026
25	1	37	1109	0.03336339044183
sum		1109		1.00
21	2	5	1317	0.00379650721336
22	2	148	1317	0.11237661351556
...	...	...	...	...

3.3.4. Estimator calculation

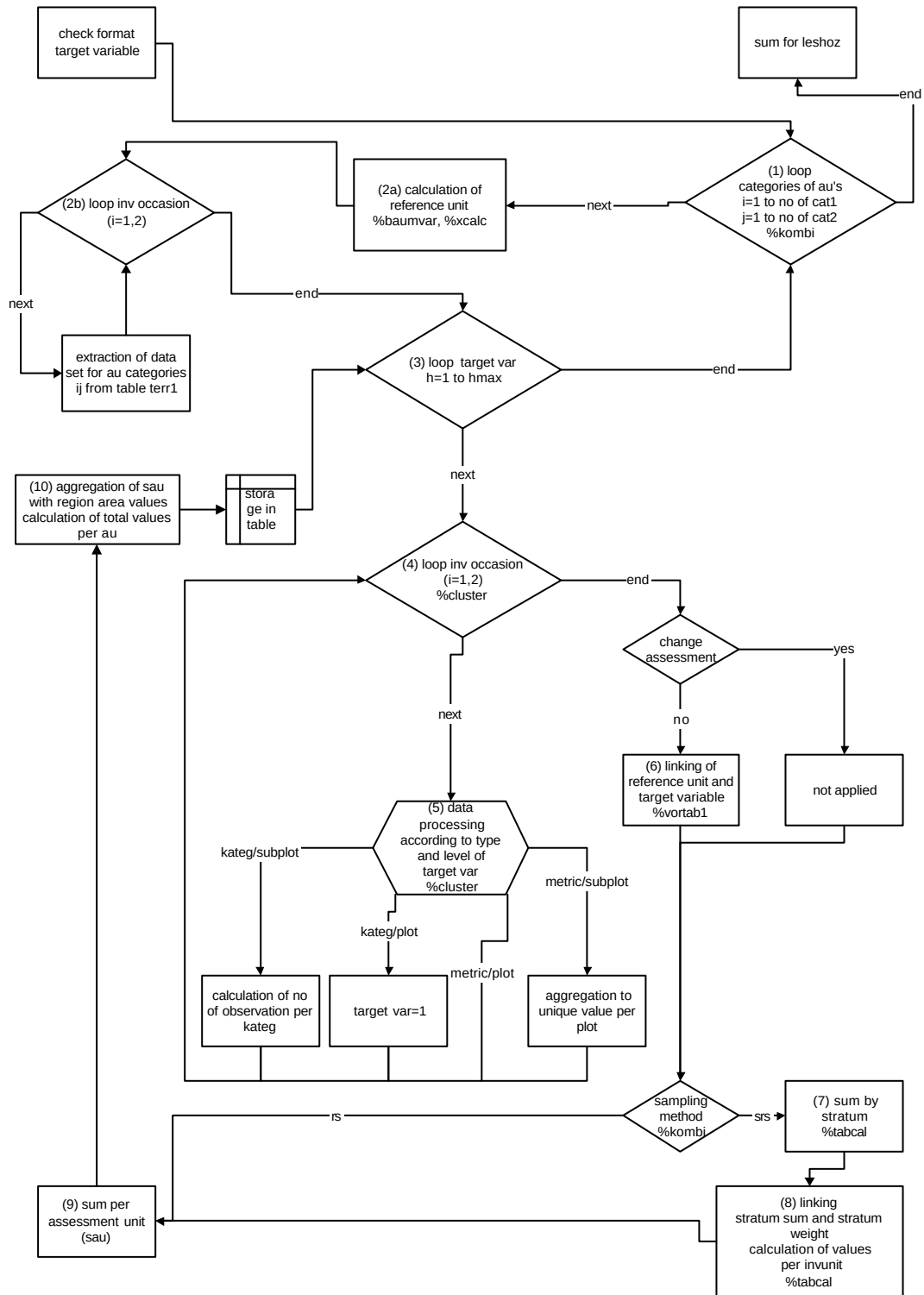


Fig. 3.3: flow chart estimator calculation. Detailed description of reference unit calculation is given in Fig. 5.2

Description of functions and data flow for estimator calculation (c.f. flowchart Fig. 3.3)

```
function kombi: controlling the estimator calculation / loop over codes of assessment units

(1) loop assessment units
begin kombi
    do i = 1 to m          /*---Loops over no of categories of assessment units---*/
        do j = 1 to n;      /*--m=no of codes au1, n=no of codes of au2---*/

            (2b) calculation of reference unit.
            %baumvar

            (2b) loop over inventory occasion
            do indi = 1 to invno      /*----Loop over no of inventories (invno=1because of state assessment in LMI---*/
                if i < m+1 and j < n+1 then do; /*---splitting data table terr1 in partitions of existing au
                    combinations---*/
                    create table resi1 as select * from table terr1
                        if cat1 = i
                        if cat2 = j
                    end
                end

                (3) loop target var
                do h=1 to hmax +1      /*--- Loop over no of codes of target var (if type =metric then hmax is assigned
                    to=1)---*/

                (4) loop over inventory occasion
                do indi = 1 to invno /*----Loop over no of inventories (invno=1because of state assessment in LMI---*/
                    %cluster          (5) data processing target var
                    end;
                    %vortab1          (6) linking reference unit and target var
                    %tabcal(&i, &j)    (7) (8) estimator calculation
                end
            end
        end
    end
end kombi
```

3.3.4.1.loop over assessment units

```
function baumvar: controlling reference unit calculation (c.f. Fig. 5.2)

begin baumvar
    proc sql;
        create table resi as
        select plotid unit2, stratum, cat1, cat2, wij
        from terr1
        where cat1= i and cat2 = j      /* see comment */
        order by plotid, unit2, stratum

        %xcalc          /*---function to calculate reference unit;---*/
    end baumvar;
```

comment on where clause: depends on allowed type of reference unit table (Fig. 3.1), only if separate reference units for au codes is allowed it is implemented in the where clause, thus the where clause could include two, one or no element.

This holds true only for the reference unit but not for the target variable which is separately calculated for each combination of assessment units irrespective of its level.

function **xcalc**: calculation of reference unit

```
begin xcalc;
  create table tempx as select * from resi (function baumvar)
  ref=1                                     /*---help var for more logical structure---*/
  xhi = max(ref)                           /*---if reference unit level = plot (e.g. forest area in
  ha)---*/
  xhi = sum(ref) weighted by wij           /*---if reference unit level = subplot (e.g. no of stems)---*/
                                          (calculates the weighted sum of subplot elements per plot,
                                          i.e. aggregation of subplot elements to plot value)---*/
  by plotid unit2 stratum
end xcalc
```

table *tempx* reference unit level = plot (e.g. forest area in ha)

plotid	unit2	STRATUM	_FREQ_	XHI
6	1	23	8	1
22	1	23	14	1
26	1	23	10	1

table *tempx* reference unit level = subplot (e.g. no of stems)

plotid	unit2	STRATUM	_FREQ_	XHI
6	1	23	8	250
22	1	23	14	640
26	1	23	10	350

function **cluster**: aggregation of subplot data to plot level; results are stored in table *vers1*

```
begin cluster
  (1) metric data
  input data set = resi1, i.e. terr1 separated by combinations of au's, i.e. all combinations of au1 and au2, all
  comb of au1 with fixed au2 and vice versa to calculate row and columns totals and one times without
  separation to calculate the grand total

  (a) target var level = sub and type = metric;
  yh1 = sum(target var) by plotid unit2 stratum weight wij, stored in table sump

  (a) target var level = plot and type = metric;
  yh1=1 by plotid unit2 stratum weight wij, stored in table sump

  (2) categorical data
  input data set = table hresi
  Building a subset of the resi1 data concerning the target var. This is a second subset of the terr1 data which
  is in addition to the au categories also separated by categories of the target var

  (a) target var level = sub and type = categorical
  %kateg
  yh1 = sum(target) by plotid unit2 stratum weight wij, stored in table vers1

  (b) target var level = plot and type = categorical
  %kateg
  yh1=1 by plotid unit2 stratum stored in table vers1
end cluster;
```

function **kateg**; results are stored in table *hresi*

begin **kateg**

```

        create table hresi from table resi1
        if catz = h      /*---category of target var---*/
        target=1        /*---default setting---*/

```

end **kateg**

function **tabcal**: estimator calculation for stratified random sampling (rs), input data set = table *vers*

begin **tabcal**

(1) aggregation level: plot

calculation of equation fragments, sum of individual observations per inventory unit and stratum

```

        suyhi suyhi2 suxhi suxhi2 suyxhi = sum (yhi yhi2 xhi xhi2 yxhi from table vers)
        by unit2 stratum, store in table tempco;

```

```

        contents table tempco:      stratum, unit2, suyhi, suyhi2, suxhi, suxhi2, suyxhi

```

```

        contents table nterr:      stratum, unit2, nht, nt, nhnt ( t indicates origin from terrestrial plots)

```

```

        contents table nhn:      stratum, unit2, nh, n, nhn

```

(2) aggregation level: stratum (point 7 in Fig. 3.3)

calculation of means variances and covariances, applies to metric and categorical target variables. If table *nha* in LMI is not available then stratification information is derived from inventory data (table *nhterr*), i.e. *a* is substituted by *nt*, *ah* by *nht*, *ahn* by *nhnt*

step1: estimator calculation

```

        create table soco as select * from tempco nterr nha      /*see ch. 6 for example */

        wh = ahn      /*resp. nhnt */
        varyh =(suyhi2-suyhi**2/nht)/(nht-1)      /*variance of y in stratum h (3.10)*/
        varxh =(suxhi2-suxhi**2/nht)/(nht-1)      /*variance of x in stratum h */
        myh  =suyhi/nht      /*mean of y in stratum h (3.6, 3.7)*/
        mxh  =suxhi/nht      /*mean of x in stratum h*/
        covyxh = (suyxhi-suyhi*suxhi/nht)/(nht-1)      /*covariance of y and x in stratum h*/
        ysth  = myh*wh      /*weighted mean of y in stratum h (3.6, 3.7)*/
        xsth  = mxh*wh      /*weighted mean of x in stratum h*/
        vy = (w**2)*varyh/nht      /*component of stratified variance estimator (3.8)*/
        vx = (w**2)*varxh/nht      /*component of stratified variance estimator*/
        r = yh/xh
        varrh=vy + r**2 * vx - 2*r*covyxh / nht*mxh**2 (Cochran, 1977,p. 166, eq 6.51)
        where clause:
        linking by key variables unit2 and stratum

```

step 2: building sums per inventory units (aggregation of stratum values to single value per invunit)

```

        y x vy vx nt varrh= sum(ysth xsth vx vy nht varrh from table soco) by unit2,
        stored in table sumvar

```

(3) aggregation level inventory unit (point 8 in Fig. 3.3)

```

        create table tmptotal as select * from sumvar, areafig(1);
        gy      = area*y
        gvy     = (area**2)*vy
        sgy     = sqrt(gvy)
        psgy    = sgy*100/gy
        gx      = area*x

```

```

gvx  = (area**2)*vx
sgx  = sqrt(gvx)
psgx = sgx*100/gx
r     = gy/gx
syx  = covxyh
gsyx = (area**2)*syx
sr    = sqrt(varr)
psr   = sr*100/r
..varr2 = (gvx+r**2*gvx-2*r**2*gsyx)/nt*gx**2      /* 3.21, Köhl and Scheuber */
where clause:
linking by key variables unit2

```

the results are to be stored cumulative in table **total** (append mode)

(1) table areafig: var *area* can be derived from table *stratarea* (sum ah) group by unit2 or if not available it is to be created from GIS data sources etc.

end **tabcal**;

function **vortab1**: aggregation/merge of data query table *terr1* and reference unit table *tempx*

```

begin vortab1;
  create table vers from vers1, tempx
  merge by plotid unit2 stratum
  yhi  =yhi;
  yhi2 =yhi**2
  xhi2 =xhi**2
  yxhi =xhi*yhi
end vortab1

```

3.3.4.2.summarizing on Leskhoz level

function **sumles**: results for Leskhoz (aggregation of inventory units) building sum per assessment unit and category of target var if type is categorical (aggregation of inventory unit values)

```

begin sumleskhoz;

calculation of equation fragments, sum of individual observations per inventory unit and categories of target
and assessment unit variables

gy gx vy vx syx nt x y gvy gvx gsyx area = sum (gy gx vy vx syx nt x y gvy gvx
gsyx area) from table total by invnr catz cat1 cat2, store in table sumles

create table tmples as select * from sumles
unit2 = 99
r      = gy/gx
varr   = (gvy + r**2*gvx - 2*r*gsyx/nt)/(gx**2)
sr     = sqrt(varr)
psr    = abs(sr*100/r)
sgy    = sqrt(gvy)
psgy   = abs(sgy*100/gy)
sgx    = sqrt(gvx)
psgx   = abs(sgx*100/gx)

append to total
end sumleskhoz

```

4. References

- Cochran, W.G., 1977: **Sampling techniques**, 3rd edition. John Wiley & Sons, New York. 428 p..
- Köhl, M., Scheuber, M., 1999: Appendix B: **Statistical estimators for the sampling design of the national forest inventory of the Kyrgyz Republic and Leshoz inventories**, In: Köhl, M., Scheuber, M., Traub, B.: Report of the Mission September 1999.
- Scheuber, M. 2000: **Leshoz Management Inventory, field manual**.
- Traub, B. 1999: Appendix C: **Proposals for a database design and data analysis concept of the national forest inventory of the Kyrgyz republic and Leshoz inventories**, In: Köhl, M., Scheuber, M., Traub, B.: Report of the Mission September 1999.

5. Appendix 1 flow charts

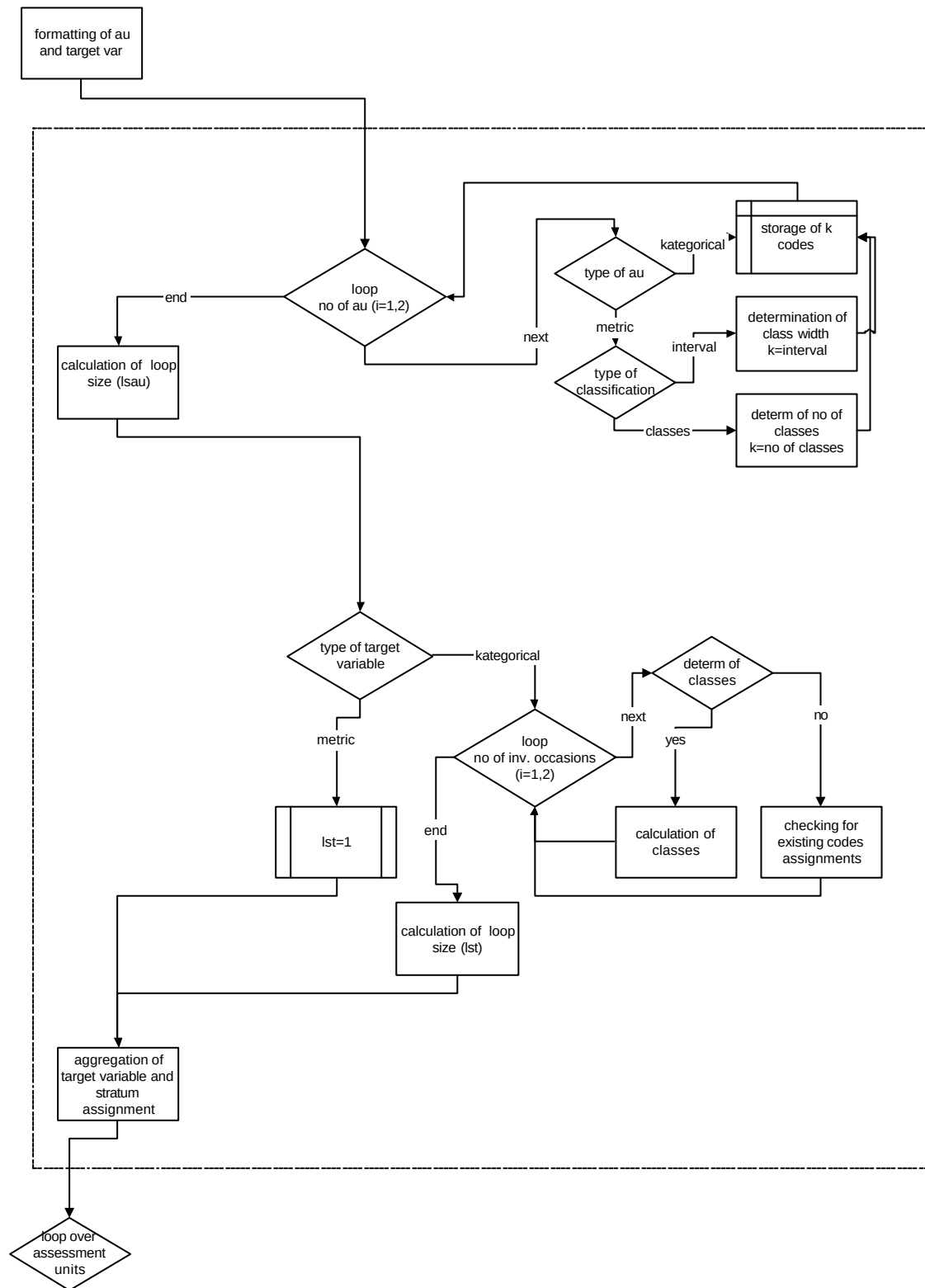


Fig. 5.1: formatting target and assessment unit variable (Detailed description of step 5 in Fig. 3.2)

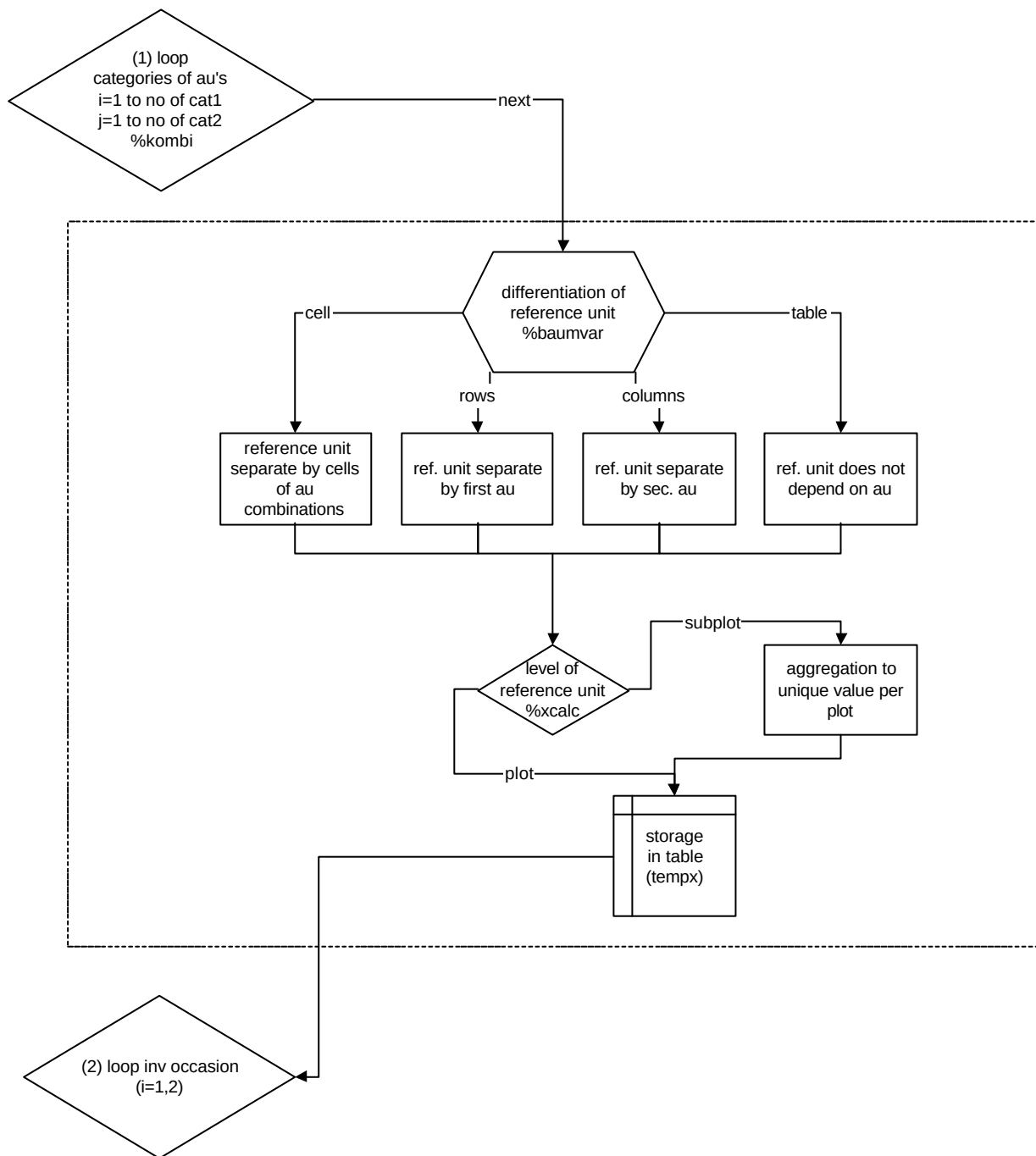


Fig. 5.2: calculation of reference unit, detailed description of point 2a in Fig. 3.3

6. Appendix 2

6.1. Sample size

For the sake of simplicity in the following, simple random sample and a population of N single trees is regarded. For the simple random sampling without replacement each arranged sequence of n selected elements has the probability $(N-n)!/N!$. From this follows for $p(s)$

$$p(s) = \begin{cases} 1/\binom{N}{n} & \text{if } s \text{ has } n \text{ elements} \\ 0 & \text{in all other cases} \end{cases}$$

For the n selected elements various characteristics can be assessed, e.g. for n trees the attributes tree species, single tree timber volume or crown transparency. If the population is defined without consideration of the values of the single tree attributes, it has size N . Likewise the selected random sample has size n , with which $p(s)$ is independent of the considered attribute and its value. That is, the random sample design $p(\bullet)$ is constant.

The meaning of this shall to be clarified by an example. With a random sample the timber volume of spruce and pine are to be estimated. Therefore from a population of N single trees a random sample is selected with size n . The attribute value of the "pine volume" is equal to the single tree volume of pines, for all other tree species equal to zero. The attribute value of the "spruce volume" is equal to the single tree volume of spruces, for all other tree species equal to zero. With this assumption both, N and n remain the same independently of the regarded attribute, equally $p(s)$ is constant.

If, however, during the derivation of estimated values only those elements are considered, for which an attribute value has been assigned (i.e. not equal to zero), N and n change depending to the considered attribute. Therewith the random sample design $p(\bullet)$ and thus $p(s)$ are depending on the regarded attribute and are no longer constant. This case can be realised also with random sample inventories, leads however to substantially more complex estimation algorithms, since for each attribute the selection probabilities must be adapted according to N and n . The following implementations therefore refer to the assumption that both N and n are constant and attributes can take the value zero.

6.2. Definitions and Notation

The values of a characteristic of the elements, which form the population, are marked with $Y_1, Y_2, \dots, Y_N$. N designates the number of elements of the population. By a random sample n elements of the N elements of the population are selected with well-known probability, These are marked with $y_1, y_2, \dots, y_n$. Based on these n elements parameters for the population are deduced. The four most important parameters are:

1. the average value, $\bar{y}$ (e.g. the mean timber volume per random sample in the inventory area in [m^3])
2. the total value, Y , (e.g. the entire timber volume in the inventory area in [m^3])
3. the proportion (Ratio) of two average values or total values, $R = \bar{y} / \bar{x} = Y/X$ (e.g. the proportion of the total timber volume to the forest area in [m^3/ha])
4. the portion of units, which fall into the same class (e.g. portion of a tree species)

For characteristics of a population capital letters are used, small letters designate the characteristics of a random sample. The symbol "^(^)" (head) stands for an estimated value of a population, which is based on random sample values.

	Population	Estimated Value
Mean	$\bar{Y}$	$\hat{\bar{Y}} = \bar{y} = \frac{\sum y_i}{n}$
Total	Y	$\hat{Y} = N\hat{\bar{Y}} = N\bar{y}$
Ratio	R	$\hat{R} = \frac{\hat{\bar{Y}}}{\hat{\bar{X}}} = \frac{\bar{y}}{\bar{x}}$
Portion	P	$\hat{p} = p = \frac{a}{n}$

general formulas for variances and covariances transformed for better computing

$$v(x) = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 / n}{n} \quad \text{eq. 6.1}$$

$$\text{cov}(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n} = \frac{\sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n}}{n}$$

where:

- a designates the number of elements in a class,
- n the sample size,
- y_i the value of an individual characteristic and
- N the total extent of the population.

Proportional stratum allocation

$$n_h = \frac{N_h}{\sum_{h=1}^L N_h} * n = \frac{N_h}{N} * n \quad (4.2)$$

Where

- L = Number of strata
- N = Number of elements in the population = $\sum N_h$
- N_h = Number of population elements in stratum h
- n = Sample size = $\sum n_h$
- n_h = Sample size in stratum h

$$v(p_{st}) = \sum_{h=1}^L w_h^2 \frac{p_h q_h}{n_h - 1} \approx \frac{1}{n} \sum_{h=1}^L w_h p_h q_h$$

$$\text{proof: } \sum_{h=1}^L w_h^2 \frac{p_h q_h}{n_h - 1} \approx \sum_{h=1}^L \frac{n_h^2}{n^2} \frac{p_h q_h}{n_h} \approx \sum_{h=1}^L \frac{n_h}{n} p_h q_h \frac{1}{n} \approx \sum_{h=1}^L w_h p_h q_h \frac{1}{n} \approx \frac{1}{n} \sum_{h=1}^L w_h p_h q_h$$

with:

$$q_h = 1 - p_h$$

6.3. Export file

Tab. 6.1: Structure of the export files.

Variable	Definition
<i>invnr</i>	Inventory identification number
<i>unit</i>	inventory Unit
<i>au1</i>	1. Assessment unit with formatted values.
<i>au2</i>	2. Assessment unit with formatted values.
<i>target</i>	Formatted target variable for categorical data.
<i>gy</i>	Estimator total of the target variable.
<i>sgy</i>	standard error for the target variable absolute.
<i>psgy</i>	standard error for the target variable percentage.
<i>gx</i>	Estimator total of reference unit.
<i>sgx</i>	standard error for the reference unit absolute.
<i>psgx</i>	standard error for the reference unit percentage.
<i>r</i>	Ratio estimator.
<i>sgr</i>	standard error of the ratio absolute.
<i>psgr</i>	standard error of the ratio percentage.

6.4. temporary tables

table *soco* (example)

unit2	STRATUM	SUYHI	SUYHI2	SUXHI	SUXHI2	SUYXHI	NHT	NT	NHNT	AH	N	AHN	VARYH	VARXH	MYH	MXH	COVYXH	YSTH	XSTH	wh
1	11	.	.	.	.	.	.	.	.	11677	19751	0.59121057 16	.	.	.	.	.	.	.	0.59121057 16
1	21	3057.4	1445526.38 92	10	10	3057.4	10	1109	0.00901713 255184	18	19751	0.00091134 63	5675.07681 33	0	305.74	1	0	0.27863500 58	0.00091134 63	0.00091134 63
1	22	36249.01	11625410.4 58	193	193	36249.01	195	1109	0.17583408 476104	1164	19751	0.05893372 49	129.183094 86	0.00005232 59	185.892358 97	0.98974358 97	1.91641607 19	10.9553291 4	0.05832927 64	0.05893372 49
1	23	191001.206	87939747.6 06	537	537	191001.206	541	1109	0.48782687 1055	4331	19751	0.21928003 65	70.1936184 7	0.00001359 08	353.052136 78	0.99260628 47	2.61520101 32	77.4172854 24	0.21765874 23	0.21928003 65
1	24	143005.965	76319199.1 78	326	326	143005.965	326	1109	0.29395852 119026	2403	19751	0.12166472 58	128.239472 45	0	438.668604 29	1	0	53.3704954 75	0.12166472 58	0.12166472 58
1	25	19697.034	13509451.5 63	37	37	19697.034	37	1109	0.03336339 044183	114	19751	0.00577185 97	2270.03813 45	0	532.352270 27	1	0	3.07266258 98	0.00577185 97	0.00577185 97
1	41	.	.	.	.	.	.	.	.	22	19751	0.00111386 77	.	.	.	.	.	.	.	0.00111386 77

remark that total leshoz area resp. sum of inv units has a stratum and area assignment, however only forest strata where field plots are established contain data for x y etc.

7. Appendix 3 data base tables

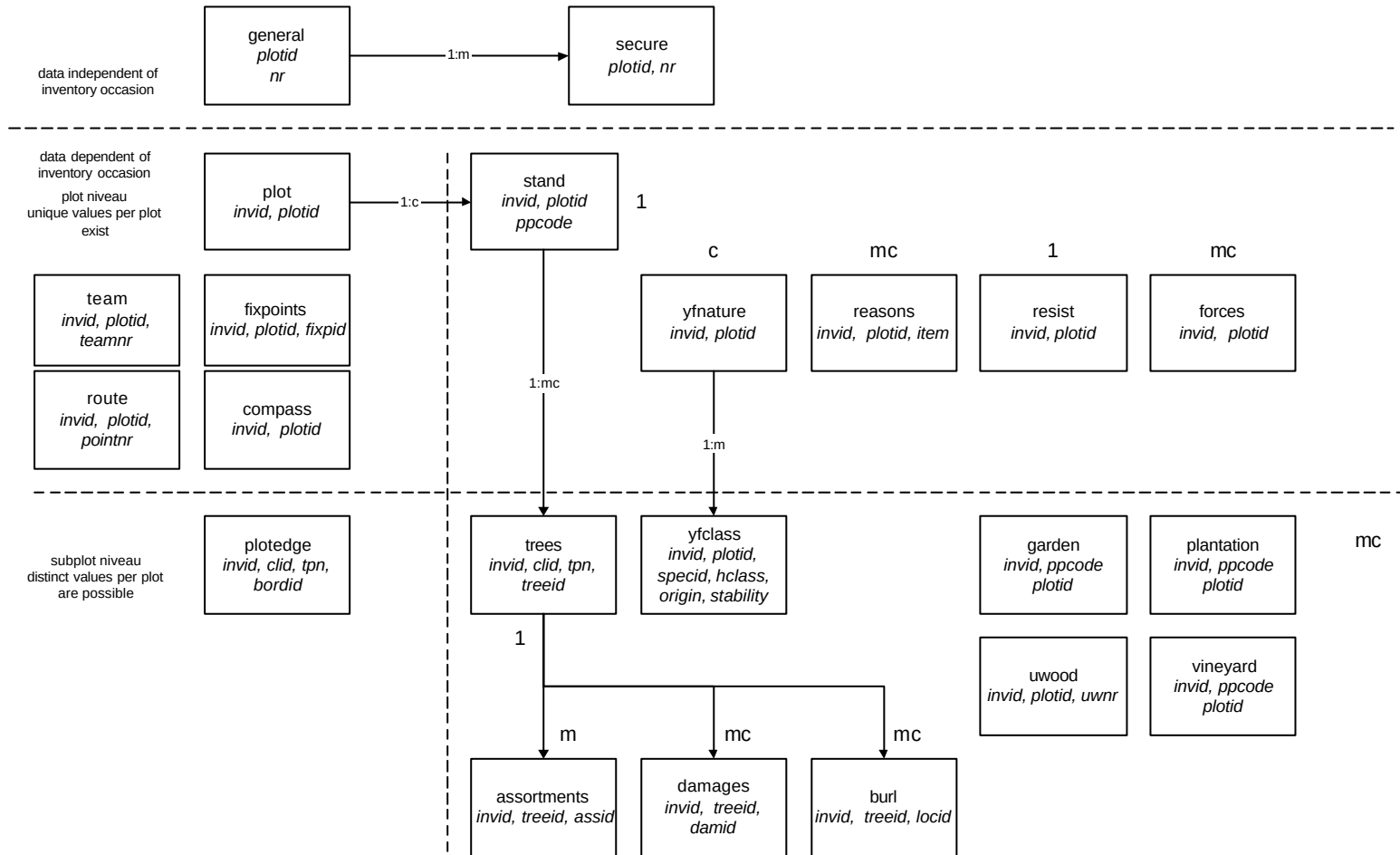


Fig. 7.1: ER diagram of LMI database

ASSORT.db

Field #	FIELD TITTLE	FIELD TYPE	KEY	POS	DESCRIPTION
	TreeID	figure	p		Tree identify code
	InvID	figure	p		Inventory number identify code
	PlotID	figure	f		Trial plot identifier
	SortKey	code	p	deriv ed	Code of assortment
	SortVol	m ³		deriv ed	Volume of assortment

ATD.db

Look-up table to select name and code for province and district.

Auto-select after selecting forestry unit.

Field #	FIELD TITTLE	FIELD TYPE	KEY	POS	DESCRIPTION
1	ATD1Code	Smallint			Province code
2	ATD2Code	Smallint			District code
3	ATDName	String(32)		2.1 2.2	Name

Burl.DB

Description of burls

Field #	FIELD TITTLE	FIELD TYPE	Key	POS	DESCRIPTION
1	PlotID	Integer	*		Trial plot identifier
2	InvID	Integer	*		Inventory number identify code
3	TreeID	String(16)	*		Tree identify code
4	Root	Smallint		4.14	Code of presence on root
5	Stem	Smallint		4.14	Code of amount on stem

Compass.DB

unit of azimuth measurements

Field #	FIELD TITTLE	FIELD TYPE	Key	POS	DESCRIPTION
1	PlotID	Integer	*		Trial plot identifier
2	InvID	Integer	*		Inventory number identify code
3	FPList	String(1)			unit of fix points azimuths
4	Route	String(1)		1.14	Unit of azimuths on route to trial plot
5	Border	String(1)		1.18	Unit of azimuths of trial plot edges
6	SecP	String(1)		1.18	Unit of securing points azimuths
7	Exp	String(1)		2.8	Unit of slope exposure
8	TreeAzi	String(1)		4.4	Unit of tally trees azimuth

Datentry.db

Actions to change database entries

Field #	FIELD TITTLE	FIELD TYPE	Key	POS	DESCRIPTION
1	Cnt	AutoInc	*		Records counter, auto
2	PlotID	Integer	*		Trial plot identifier
3	InvID	Integer	*		Inventory number identify code
4	Date	Date	*		Date of access to database
5	Action	Smallint			Code of database changing
6	Operator	String(32)			Name of operator
7	Note	String(255)			Remarks

Fixpoint.db

List of fix points

Field #	FIELD TITTLE	FIELD TYPE	Key	POS	DESCRIPTION
1	PlotID	Integer	p		Trial plot identifier
2	InvID	Integer	p		Inventory number identify code
3	FPNr	Smallint	f		Fix point number
4	FPX	Integer			X coordinate of fix point
5	FPY	Integer			Y coordinate of fix point
6	TPAlt	Integer			Altitude of trial plot (from the map)
7	Azi	Float			Azimuth from fix-point to trial plot centre (on map)
8	MapDist	Integer			Distance from fix-point to plot centre (on map)
9	FPAAlt	Integer			Altitude of fixed point (from the map)
10	FPDescr	String(255)			Description of the fix point
11	FldDist	Integer			Distance from fix-point to plot centre (for the field)
12	AziG	Integer			Azimuth from fix-point to trial plot centre in gons

Forces.db

Forces or pressures, that influence the forest stand

Field #	Field Tittle	Field Type	Key	POS	DESCRIPTION
1	PlotID	Integer	*		Trial plot identifier
2	InvID	Integer	*		Inventory number identify code
3	Type	Smallint	*	3.16	Code of force type
4	Degree	Smallint		3.16	Degree of force influence

Forestry.DB

Look-up table to select code and name of forestry units (Leskhoz, Lesnichestvo).

Codes of administrative unit use to select province and district

Field #	Field Title	Field Type	Key	POS	DESCRIPTION
1	F1Code	Smallint			Code for forestry (Leskhoz)
2	F2Code	Smallint			Code for forestry farm (Lesnichestvo)
3	Fname	String(64)		2.4 2.5	Name of forestry unit
4	ATD1Code	Smallint			Code for province
5	ATD2Code	Smallint			Code for district

Garden.db

Plantations of fruit trees or shrubs

Field #	Field Title	Field Type	Key	POS	DESCRIPTION
1	PlotID	Integer	*		Trial plot identifier
2	InvID	Integer	*		Inventory number identify code
3	IntNr	Smallint	*		Intrinsic to form number
4	Spec	Smallint		6.1	Code for specie
5	Age	Smallint		6.1	Age of garden
6	Quant	Smallint		6.1	Quantity of trees

General.DB

General & geographical plot information

Field #	Field Title	Field Type	Key	POS	DESCRIPTION
1	PlotID	Integer	*		Trial plot identifier
2	X	Float		1.8	X trial plot coordinate
3	Y	Float		1.8	Y trial plot coordinate
4	Prov	Smallint			Code for province (oblast) name
5	Distr	Smallint			Code for district (raion) name
6	Settl	String(32)		2.3	Name of settlement
7	Owner	String(1)		2.4	Code for ownership
8	Unit1	Smallint		2.5?	Code for forestry ("Leskhoz")
9	Unit2	Smallint		2.6?	Code for forest area ("Lesnichestvo")
10	Slope1	Smallint		2.7	Slope incline 1
11	Slope2	Smallint		2.7	Slope incline 2
12	Slope	Float		2.7	Average
13	Exp	Float		2.8	Slope Exposure
14	ExtRad	Float		4.2	External Circle Radius
15	IntRad	Float		4.2	Internal Circle Radius
16	Relief	String(1)		2.9	Terrain relief
17	MRelief	String(1)		2.10	Trial plot micro relief
18	ExpG	Float			Slope exposure in gons

Increment.db

Increment information

Field #	Field Title	Field Type	Key	POS	DESCRIPTION
1	PlotID	Integer	*		Trial plot identifier
2	InvID	Integer	*		Inventory number identify code
3	TreeNr	Smallint	*	3.20	Tree number
4	Spec	Smallint		3.20	Specie
5	DBH	Float		3.20	Diameter
6	RingNr	Smallint		3.20	Number of rings
7	Length	Float		3.20	Length
8	Age	Smallint		3.20	Age of tree

Plantat.db

Industrial plantations

Field #	Field Title	Field Type	Key	POS	DESCRIPTION
1	PlotID	Integer	*		Trial plot identifier
2	InvID	Integer	*		Inventory number identify code
3	IntNr	Smallint	*	6.3	Intrinsic to form number
4	Spec	Smallint		6.3	Code for specie
5	Age	Smallint		6.3	Age of trees
6	Quant	Smallint		6.3	Quantity

Plot.DB

Plot information

Field #	Field Title	Field Type	Key	POS	DESCRIPTION
1	PlotID	Integer	*		Trial plot identifier
2	InvID	Integer	*		Inventory number identify code
3	PlotNr	Integer	*	1.1	Trial plot number from field form
4	Date	Date		1.2	Date
5	Start	Time		1.3	Working time: start
6	End	Time		1.3	Working time: end
7	MapNr	String(16)		1.6	Map number
8	MapScl	String(3)		1.6	Map scale
9	Misld	String(9)		1.7	Misleading
10	TPAltM	Smallint		1.9	Altitude of trial plot from map
11	TPAltF	Smallint		1.9	Altitude of trial plot from field
12	FPNr	Smallint		1.10	Number of used fix point
13	FixX	Integer		1.11	X fix point coordinate
14	FixY	Integer		1.11	Y fix point coordinate
15	FPDesc	Memo(240)		1.13	Fix point description
16	FPAItM	Smallint		1.12	Altitude of fix point from map

17	FPAltF	Smallint		1.12	Altitude of fix point from field
18	Access	String(1)		1.15	Accessibility of trial plot space
19	LandUse	String(1)		1.16	Land use
20	Border	String(1)		1.17	Borders (Yes/No switch)
21	Route	Memo(240)		1.21	Description of the route to the fix point
22	StartAlt	Smallint		1.21	Altitude at start
23	FlightNr	String(16)			For Trial plot list
24	SqNr	String(16)			Square number
25	PhotoNr	String(16)			Aerial photo Number
26	IsForest	Smallint			“Forest / non forest” decision
27	AltDem	Smallint			Altitude from digital elevation model
28	luc			d	land use class
29	stratum			d	stratum assignment (10*luc+stgroup)

Plotedge.db

Border line of forest stock or accessibility in trial plot area

Field #	Field Title	Field Type	Key	POS	DESCRIPTION
1	PlotID	Integer	*		Trial plot identifier
2	InvID	Integer	*		Inventory number identify code
3	Dist	Float		1.18	Distance from trial plot centre to edge point
4	Azi1	Float		1.18	Azimuth 1 of border line from the edge point
5	Azi2	Float		1.18	Azimuth 2 of border line from the edge point
6	AziE	Float		1.18	Azimuth from the edge point to the trial plot centre
7	Azi1G	Float			Azimuth 1 in gons
	Azi2G	Float			Azimuth 2 in gons
9	AziEG	Float			Edge's azimuth in gons

PTID.db

One-record table to store trial plot ID and tree ID.

FIELD #	FIELD TITLE	FIELD TYPE	KEY	POS	DESCRIPTION
1	PlotID	Integer			Trial plot identify code
2	TreeID	String(16)			Tree identify code

Reasons.DB

Description of reasons of n on accessibility or non-forest decision

FIELD #	FIELD TITLE	FIELD TYPE	KEY	POS	DESCRIPTION
1	PlotID	Integer	*		Trial plot identifier
2	InvID	Integer	*		Inventory number identify code
3	Item	String(4)	*	1.15 1.16	Number of item in field form to fill out the reason
4	Reason	String(255)		1.15 1.16	Definition

Resist.DB

Resistance of the forest stand

FIELD #	FIELD TITLE	FIELD TYPE	KEY	POS	DESCRIPTION
1	PlotID	Integer	*		Trial plot identifier
2	InvID	Integer	*		Inventory number identify code
3	Spec	Smallint		3.16	Species of the stand
4	H_D	Smallint		3.16	Height / Diameter relation
5	CrownL	Smallint		3.16	Crown length
6	CrownF	Smallint		3.16	Crown form
7	VertAx	Smallint		3.16	Vertical axis
8	Roots	Smallint		3.16	Roots
9	Height	Smallint		3.16	Height
10	Gap	Smallint		3.16	Gaps in the stand
11	Struct	Smallint		3.16	Structure of the stand
12	SanStat	Smallint		3.16	Sanitary status of the stand

Route.DB

polygon steps from fix point to trial plot centre

FIELD #	FIELD TITLE	FIELD TYPE	KEY	POS	DESCRIPTION
1	PlotID	Integer	*		Trial plot identifier
2	InvID	Integer	*		Inventory number identify code
3	PointNr	Smallint	*	1.14	Point number
4	Alt	Smallint		1.14	Altitude
5	Azi	Float		1.14	Azimuth
6	DistM	Float		1.14	Distance on map
7	DistF	Float		1.14	Field distance
8	AziG	Smallint			Azimuth in gons

Secure.db

Points to mark trial plot centre

FIELD #	FIELD TITLE	FIELD TYPE	KEY	POS	DESCRIPTION
1	PlotID	Integer	*		Trial plot identifier
2	InvID	Integer	*		Inventory number identify code
3	Nr	Integer	*	1.19	Securing point number
4	Azi	Float		1.19	Azimuth of securing point
5	Dist	Float		1.19	Distance from point to trial plot centre
6	Note	String(64)		1.19	Characterisation
7	AziG	Smallint			Azimuth in gons

Stand.DB

Stand information

FIELD #	FIELD TITLE	FIELD TYPE	KEY	POS	DESCRIPTION
1	PlotID	Integer	*		Trial plot identifier
2	InvID	Integer	*		Inventory number identify code
3	Eros	String(1)		2.11	Traces of erosion (Yes/No switch)
4	GrLoc	String(1)		3.1	Growth location
5	Origin	String(1)		3.2	Origin
6	DevelSt	Smallint		3.3	Development stage
7	Type	Smallint		3.4	Forest type
8	Density	Smallint		3.5	Density
9	WalnutP	Smallint		3.6	Walnut Portion
10	Shrubs	String(1)		3.7	Kind of shrubs
11	PPlant	String(1)		3.8	Purposeful plantation
12	Age	Smallint		3.9	Age
13	Struct	String(1)		3.10	Structure
14	UnderW	String(1)		3.11	Underwood
15	StandBrd	String(1)		3.13	Stand borders (Yes/No switch)
16	Disorder	Float		3.14	Sanitary status: disorder degree
17	DeadW	Float		3.14	Sanitary status: deadwood
18	Dim1	Float		3.14	Sanitary status: Area – dimension1
19	Dim2	Float		3.14	Sanitary status: Area – dimension2
20	Pests	String(1)		3.14	Sanitary status: forest pests (Yes/No switch)
21	Disease	String(1)		3.14	Sanitary status: forest diseases (Yes/No switch)
22	Viabil	String(1)		3.17	Viability
23	Grazing	String(1)		3.18	Grazing (Yes/No switch)
24	Grass	Smallint		3.19	Grass spread
25	YFType	String(1)		5.1	Nature of young forest
26	stgroup			???	stand group

Team.DB

Team members

FIELD #	FIELD TITLE	FIELD TYPE	KEY	POS	DESCRIPTION
1	PlotID	Integer	*		Trial plot identifier
2	InvID	Integer	*		Inventory number identify code
3	TeamNr	Integer	*	1.4	Team number code
4	Name1	String(32)		1.5.1	The name of the chief of the fieldwork team
5	Name2	String(32)		1.5.2	Auxiliary staff person
6	Name3	String(32)		1.5.3	Auxiliary staff person
7	Name4	String(32)		1.5.4	Localisation staff person
8	Name5	String(32)		1.5.5	Localisation staff person

TNames.DB

Look-up table to choose field worker name from a list

FIELD #	FIELD TITLE	FIELD TYPE	KEY	POS	DESCRIPTION
1	Name	String(32)			Name
2	FCode	Smallint			Forestry (Leskhoz) code

Trees.DB

Tree information

FIELD #	FIELD TITLE	FIELD TYPE	KEY	POS	DESCRIPTION
1	PlotID	Integer	*		Trial plot identifier
2	InvID	Integer	*		Inventory number identify code
3	TreeID	String(16)	*		Tree identify code
4	Number	Smallint		4.3	Internal tree number
5	Azi	Float		4.4	Azimuth
6	Dist	Float		4.5	Distance
7	Spec	Smallint		4.6	Specie
8	DBH	Float		4.7	Diameter
9	D7	Float		4.8	Diameter on 7 meters
10	Height	Float		4.9	Height
11	Layer	Smallint		4.10	Layer
12	Class	Smallint		4.11	Tree class
13	D_L	Smallint		4.12	Damage: location
14	D_C	Smallint		4.12	Damage: cause
15	MarketC	Smallint		4.13	Marketability: class
16	MarketL	Float		4.13	Marketability: length
17	Note	String(32)		4.15	Remarks
18	AziG	Smallint			Azimuth in gons

Uwood.db

Prevailing under wood species

FIELD #	FIELD TITLE	FIELD TYPE	KEY	POS	DESCRIPTION
1	PlotID	Integer	*		Trial plot identifier
2	InvID	Integer	*		Inventory number identify code
3	UWNr	Smallint	*	3.12	Underwood record number
4	Spec	Smallint		3.12	The most frequent under wood specie

Vineyard.db

Vineyard plantations

FIELD #	FIELD TITLE	FIELD TYPE	KEY	POS	DESCRIPTION
1	PlotID	Integer	*		Trial plot identifier
2	InvID	Integer	*		Inventory number identify code
3	Age	Smallint		6.2	Age of garden
4	Quant	Smallint		6.2	Quantity

Yfnature.DB

General Young forest information

FIELD #	FIELD TITLE	FIELD TYPE	KEY	POS	DESCRIPTION
1	PlotID	Integer	*		Trial plot identifier
2	InvID	Integer	*		Inventory number identify code
3	type	Smallint		5.1	nature of young forest

Yfclass.DB

Young forest classes

FIELD #	FIELD TITLE	FIELD TYPE	KEY	POS	DESCRIPTION
1	PlotID	Integer	*		Trial plot identifier
2	InvID	Integer	*		Inventory number identify code
3	Specie	Smallint	*	5.2	Specie of young forest
4	Class	Smallint	*	5.3	Height class of young forest
5	origin	Smallint	*	5.3	origin of plants (seed / veget.)
6	stability	Smallint	*	5.3	condition of plants (strong / weak)
7	number	Smallint		5.3	number of plants