4.2 Analysis Software

Berthold Traub, Hans-Jörg Schnellbächer

The inventory concept of the National Forest Inventory requires the analysis software to be highly flexible. The current value of various measures of the Swiss forests and their changes since the first inventory are of interest. A breakdown of the results into geographical units, such as production regions, cantons, or economic regions is just as important as the formation of assessment units, which allows further differentiation of the results according to categories, such as ownership or forest type.

For the design of the data storage and the analysis software, the following concept was pursued in order to permit the various analyses:

The attributes are stored in their original, non-aggregated form in a relational database (i.e., just as they were assessed) (see Chapter 4.1). The aggregation is conducted by special analysis software which ensures the correct access to the database and the right application of the analysis method. The results can be exported for the data exchange in table, graphic, or file format.

The data analyzed in the second NFI were: 1) assessed in the field by measuring the sample plots; 2) by enquiry; 3) obtained through interpretation of aerial photographs; 4) extracted using GIS analysis; or 5) generated by models. The analysis software can be interpreted as a "data warehouse" (MATTISON 1997). The software links data from different sources and information levels and provides aggregated or multi-dimensional data. This ensures a consistent foundation of data for all users.

The analysis software should enable a user to analyze National Forest Inventory data in a simple and fast manner. In order for a wide user group to carry out analyses independently the software, the statistical analysis methods, and the databank management should all be easy to use without any special training.

The double sampling for stratification was implemented in SAS (Statistical Analysis System), version 6.12. To ensure user friendliness and to avoid errors of the definitions for the input variables, a graphical user interface was developed for the access of analysis routines. The SAS system offers several different options for this. The module SAS/AF (SAS Institute 1989) together with the screen-control-language (SAS Institute 1990), was used for the NFI analysis software. It is possible to put together an analysis (parameterization) with only a few steps by using predefined, partially flexible selection lists and selection fields.

The database query and the calculation of the estimators can consume considerable amounts of database resources. Compared to this, relatively few resources are necessary in order to store the definitions of the query and the results of the analysis. The parameterization, the data analysis, and their presentations are, therefore, carried out separately. The parameter and the results of the analysis are stored in SAS data files, but not the raw data itself. This way it is possible to easily reproduce the analyses at a later time or to quickly visualize the results with different types of presentations (tables, graphics). The analysis in batch mode is possible as well.

The parameter definitions are managed within projects that encompass different inventories or thematic areas. Furthermore, analysis prototypes – organized in thematic catalogs – are provided which can be adopted and altered to individual parameter definitions.

A context sensitive online help function is available for explanations of the windows and entry fields. This function contains general remarks about the adjustments, the input of parameters, the information about options and their effects. For individual analyses different SAS modules such as SAS-Assist, SAS-Insight or the Program-Interpreter from SAS are available.

The NFI analysis software is arranged into four main areas:

1) The project management for managing the analysis definitions
2) The definition of an analysis (parameterization)

3) The analysis (databank query and calculations of the estimators)
4) The presentation of the results

In order to facilitate the work some programs for the control of the system and for data management were added as well.

4.2.1 Parameterization of the Analysis

General

The organization of a correct parameterization is ensured in two different ways: First, the possible selections in lists and selection fields are adapted to the parameters set. Second, before a parameterization is recorded and stored, a list of tests are conducted which examine the validity of the parameter combination. Analyses that are methodologically incorrect are, for the most part, avoided.

The most important parameters and their effects on the analysis are presented briefly in the following. After the definition of the project, in which the desired subjects are managed or are newly included, the subject can be dealt with. An overview of all the entry fields can be found in Table 1.

Table 1. Content of the window “parameterization”.

Theme	Seven digit name of a certain data analysis (each theme is a record in a table containing the whole set of analysis parameters). A selection list that consists of theme definitions is displayed.
Description	Short description of the analysis, which is used as the title for tables and graphs.
Library	SAS-libref (library reference) defines the folder where the analysis results are stored.
Inventory perimeter	Determines the analyzed region. It is possible to select between 1) ”CH” (whole country) for the analysis of the NFI and 2) “canton” for the analysis of cantonal inventories. The other analysis options are adjusted to the selected perimeter.
Inventory cycle	It is possible to analyze the state (NFI1, NFI2) and change, or conduct an analysis on the joined grid.
Analysis unit	The analysis unit defines the underlying population (number of terrestrial sample plots) for the analysis. The analysis usually refers to the “accessible forest without shrub forest.”
inventory unit	The inventory unit is a geographically clearly defined area. The estimates are calculated separately for the individual inventory units. The overall estimate for the inventory perimeter of interest is the sum of the estimates from the individual inventory units.
Condition	Two conditional fields are available. For change analysis, the second conditions affect only the recent inventory. For state analysis, both conditions affect the selected inventory.
Target variable	From a menu list several variables can be selected or an alias can be entered for an SQL function. Options: In one of the two-option fields (1) the reference unit can be selected. In a sub-menu, it is possible to select whether the reference unit is calculated separately for each table cell or uniformly for columns, rows, or the entire table. It is also possible (2) to enter a function for changing/editing the target variable.
NVL	The two-digit input field determines the treatment of missing value or null values. By default they are replaced during the database query by the value zero.
Assessment unit	The smallest unit for the calculation of the estimators. If nothing is specified, the assessment unit and the inventory unit are identical. Several different options offer the possibility to create classes, or the integration of lookup tables etc.
Save	Saving the parameter and starts the analysis if desired.
End	Leaves the window without saving.

Inventory Cycle
Information from different inventory dates is stored in the databank. The field “inventory cycle” allows the user to select between state analysis (only one inventory) and a change analysis. The change can be calculated as the total change (difference between the two states) or as the mean annual change. Furthermore, it is possible to calculate the current values of both inventories on the joined grid. This is important for the analysis of ratio estimators. The individual options are listed in Table 2.

Table 2. Options for the inventory selection.

State	Analysis of data sampled at a certain inventory occasion. The inventory can be selected from a list.
Change	Analysis of change between two inventories. The initial state is selected from list 1; the final state is determined in list 2. Subject of the analysis is the difference between inventory 2 and inventory 1. By default, weighting and classification is carried out according to the current inventory data (difference class2). For special analysis it is possible to use both inventories for the classification (difference classes 1/2). For the classification of the difference, the option “classification difference” was created. In order to accomplish this, the difference of a continuous target variable is calculated first and the results are classified. See also: classification of the target variables. For the change analysis of proportions, the option “state / joined grid” exists. For this, a state analysis of the first and second inventory occasion is carried out and saved separately in the result table.
Reference period	The reference period determines whether changes are calculated for the entire inventory period or as annual change. This is especially important in case of varying length of inventory periods.

Analysis Unit
Selecting the analysis population determines the size of the sample. The population can refer to the entire 500-meter-grid (GIS data, air photo data) or to a sub-grid (terrestrial survey). The selection of the analysis population also effects the statistical analysis methods. Data from the entire grid are only analyzed in one phase, while data on the terrestrial grid is analyzed in two phases (double sampling) (see Table 3). The selection of the analysis population determines the target variables available for the analysis as well.

Inventory unit
The results are at first derived for distinct geographic subunits, such as production regions, economic regions, cantons, or forest districts. The total values for Switzerland or the cantons (for the cantonal analysis) are the sum of the individual results of the inventory units. The employment of different inventory units causes a slight difference between the total values. This is caused, on the one hand, by the calculation method and, on the other hand, by different total areas of the inventory unit. The total area for the cantons and the forest districts (4,128,419 hectares, derived from the municipal borders of the Swiss Federal Statistical Office GEOSTAT as of January 1, 1994) are seven hectares smaller than the area in the production regions, economic regions, and protection forest regions (4,128,426 hectares, derived from the partition of municipalities of the National Forest Inventory from 1985). The following inventory units can be selected:
Forest enterprise (only for special analysis in the cantons)
Forest district in the cantons (only for the cantonal analysis)
Forest compartment of the cantons (only for the cantonal analysis)
Canton
Production regions (standard setting)
Protection forest regions
Economic regions

Table 3. Definition of the analysis unit.

Analysis unit	Description	DS	Tables for the target variables	Condition
Total	All of the air photo samples and GIS data.	No	beo.cl, lfi2.cl2, lfi2.lbaufn, lfi2.strasse, lfi2.holzweg	none
Total forest	Aerial photography samples and GIS data for which the forest decision = forest.	No	beo.cl, lfi2.cl2, lfi2.lbaufn, lfi2.strasse, lfi2.holzweg	lbaufn. kombent = 2
Shrub forest	Aerial photography samples and GIS data for which the forest decision = shrub forest.	No	beo.cl, lfi2.cl2, lfi2.lbaufn, lfi2.strasse, lfi2.holzweg	lbaufn. kombent = 3
Terrestrial forest	Terrestrial survey without shrub forest, including inaccessible samples.	Yes	lfi2.wa, lfi2.ba, lfi2.schaden, lfi2.sortd, lfi2.bemerk, lfi2.hhvorauf, lfi2.ruvorauf, beo.ruecken (, lfi2.strasse)	lbaufn. kombent = 2
Accessible forest	Accessible terrestrial sample plots without shrub forest.	Yes	lfi2.wa, lfi2.ba, lfi2.schaden, lfi2.sortd, lfi2.bemerk, lfi2.hhvorauf, lfi2.ruvorauf, beo.ruecken (, lfi2.strasse)	lbaufn. kombent = 2 wa.zugang < 3
Wooded area	Accessible terrestrial sample plots without shrub forest, with utilization: “Forest in a strict sense.”	Yes	lfi2.wa, lfi2.ba, lfi2.schaden, lfi2.sortd, lfi2.bemerk, lfi2.hhvorauf, lfi2.ruvorauf, beo.ruecken (, lfi2.strasse)	lbaufn. kombent = 2 wa.nutzkat >8
Young growth	Terrestrial young growth survey. Each young growth plot counts as a sample plot.	Yes	lfi2.jwsalfi2, lfi2.jwkla2, beo.jwgru	lbaufn. kombent = 2 jwsalfi2.jwlage in (1,2,4)
Forest edge	All terrestrial plots that intersect a forest edge.	No	lfi2.wr, lfi2.wrartant	lbaufn. kombent = 2

Conditions

For the analysis of certain subjects (for example, analysis of the timber volume, increments, and utilization) it is, at times, necessary to restrict the data which is to be analyzed. For the restriction, two separate conditions can be entered. These restrictions must contain valid variables from the database and must be correctly linked using ORACLE SQL.

For the analysis of current values both restrictions are equally treated and used. For the change analysis the first restriction is applied to the query of data for both inventories (see Table 4). The second restriction only affects the data query of the second inventory.

Table 4. Conditions for the change analysis.

Condition 1	Restrictions that refer to the target variable (e.g., elimination of sample plots with missing stability assessment for the analysis of the stand stability). Restrictions, like tree history for the analysis of growing stock and increment etc.
	Restrictions that refer to attributes that do **not** depend on the inventory occasion).
Condition 2	Restrictions that refer to the definition of assessment units (e.g., if the increment only for certain types of forests is of interest). For change analyses the second occasion value of the variable that determines the assessment unit is principally used.
	Restrictions for attributes which definitions or coding depend on the inventory (damage, etc.).

Target variable

The term “target variable” refers to the variable to be analyzed. The analysis is adjusted depending on the type (continuous or categorical) and the level (sample plot or subunit) of a target variable. The properties of the target variables must be manually adjusted. The following options are available:

– For the manipulation of the target variable it is possible to specify an SQL expression.
– The classification of the target variable
– Avoiding several identical data sets (example tree damage analysis)
  For this the option DISTINCT should be selected.

Reference Units

The reference unit of the analysis is used for the calculation of ratio estimates (for example, volume per hectare) (KÖHL 1994). Several different reference units such as area in hectares, number of stems, timber volume, basal area, etc., or the target variable itself can be selected in order to calculate percentage values.

In the input field "weight," an SQL expression can be inserted for the weighting of the reference unit, starting with an operator (*, /, +, –). Furthermore, the type of reference unit table can be selected.

The possible options, depending on the level of the target variable and the assessment unit, are described in Figure 1.

Options:
– Reference to a cell — A reference unit is derived separately for each cell
– Reference to a column — A common reference unit is derived for each column
– Reference to a row — A common reference unit is derived for each row
– Reference to a table — A common reference unit is derived for each table

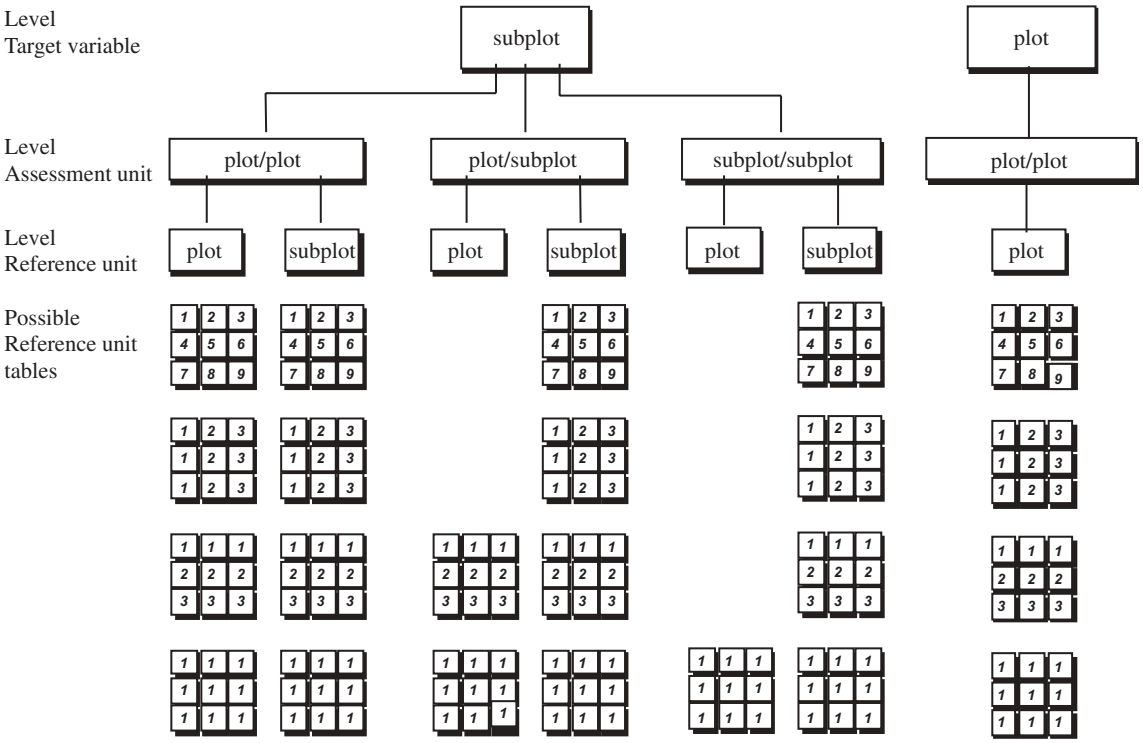

Figure 1. Schematic of possible reference unit tables.
Equal values per row or column mean that a unique value for the reference unit is applied.

Assessment Units

The analysis can be conducted separately according to categories of a certain attribute. The definition of these categories determine the number of rows and columns of the result table and thus the "thematic resolution" of the analysis. Five options are available for the manipulation of the group variables (see Table 5).

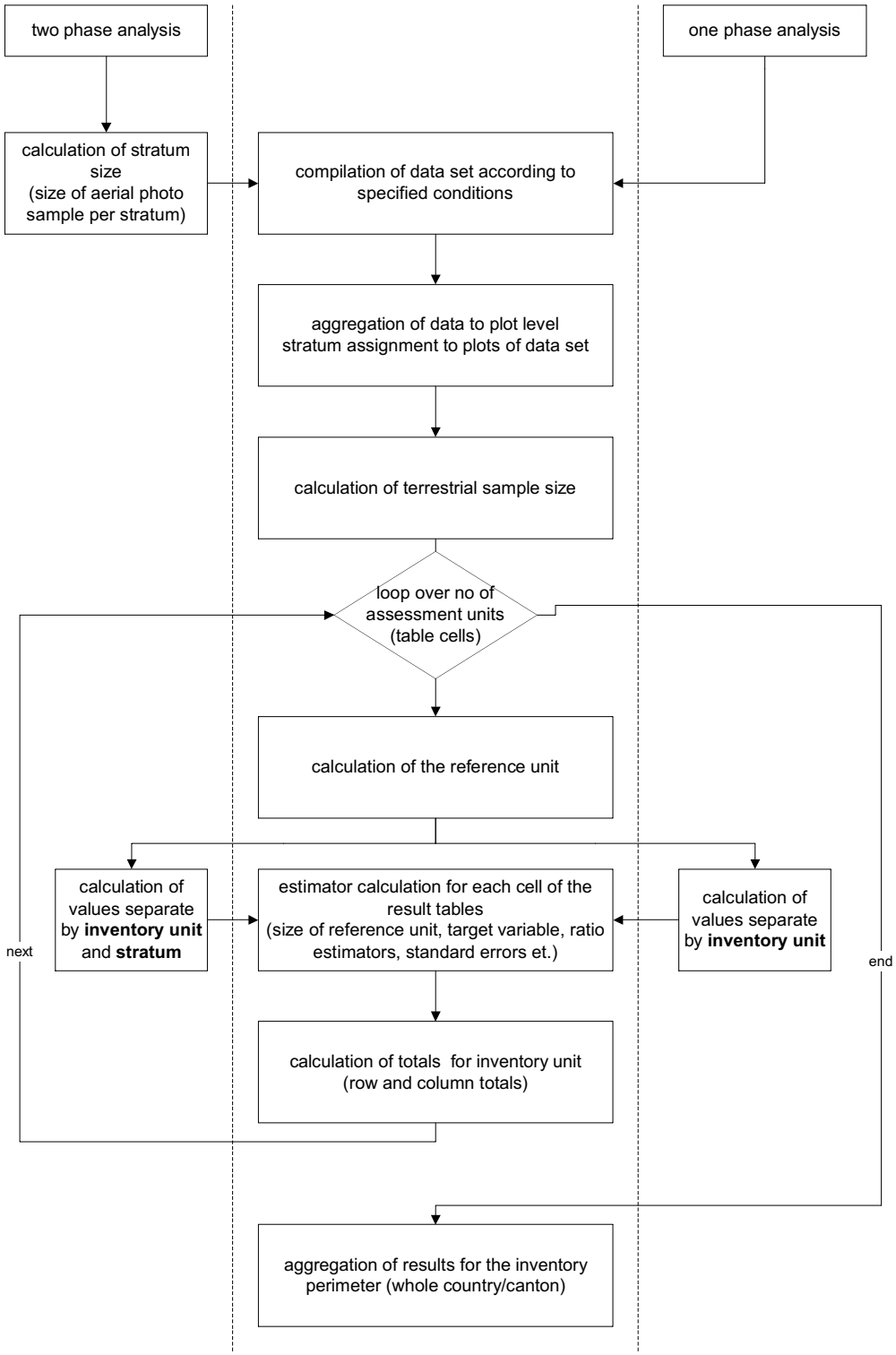

Figure 2. Flow chart of the analysis.

(Fictitious) example:
For the analysis of timber volume (level: subplot) by assessment units: 1) ownership (level: plot), and 2) tree species (level: subplot), the following several possibilities exist for the reference unit (a) hectare (level: plot) and reference unit (b) number of stems (level: subplot):

a) Reference unit hectares
   The timber volume per hectare can either be calculated for each type of ownership or for the entire forest area. 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.
b) Reference unit number of stems
   The number of stems allows for the selection of all possible table types. It is also possible to calculate the mean volume per tree according to: 1) ownership and 2) tree species. The remaining types of tables are mainly appropriate for the calculation of proportions.

Table 5. Options for the classification of assessment units.

CODE	Direct application of the codes for categorical variables
INTERVAL	For simple classification of continuous variables for an arbitrary number of classes.
CLASSES	For the classification of continuous variables in classes with a fixed reference point and class width or any classification by entering the class limits. For the classification two options area available: (1) Classifications with a fixed width for any reference point. Class width and starting point are defined. (2) Classification with arbitrary width. The desired number of classes and the class limits are defined. The classification table is stored in the result library under the corresponding variable name. For both types of classification, it is possible to treat the class limits differently. Either the lower or the upper limit (default) can be included in the class.
OTHERS	For arranging variables of any arbitrary scale (e.g., dividing ownership category in (1) public (2) and private. A key with the created code is generated and can be edited.
LOOKUP	With this option it is easy for the user to carry out complex manipulations of the assessment unit classification. It is possible to use additional information (e.g., from the GIS) in an analysis to change the original codes. The tables are linked by the cluster identification number. The following elements can be found in the window “look-up table”: OWNER: Selection list of the owner (ORACLE). TABLE: Selection list of the available tables. (ORACLE). CODEVARIABLE: Selection list for variables containing the new code to be assigned. FORMAT VARIABLE: Selection list for plain text variables. TABLE INFORMATION: This displays only the description of the selected elements and the content of the selected table. JOIN-TABLE: to which the lookup table is linked to. If an alias was specified as the assessment unit, the name of the appropriate table, to which the look-up table is linked to, has to be manually adjusted with the selection keys. CONDITIONS for lookuptable: The necessary conditions for the WHERE clause to link to the look-up table has to be specified here. It is important to correctly treat the NULL values. (e.g., AND NVL(ba.bart,0) = lfi2.nhlhlut.bart)

4.2.2 Process of the Analysis

After the parameters are stored the analysis can be started directly or in a batch mode. The process of the analysis is presented in a simplified form in Figure 2. For the one-phase analysis the estimates are derived separately for the inventory units. For the double sampling method the data need to be stratified. The calculation of the estimates is carried out for the assessment units of each stratum. They are then weighted according to the stratum size and are summarized to inventory units.

In both cases the values of the inventory units add up to total values for Switzerland or for the cantons, and are stored in the results file.

4.2.3 Presentation of the Results

Analysis

The window "analysis" is used for the output of the analysis results. The presentation of the results can be individually adjusted using different selection menus. The settings for each subject are stored in a separate file. If the subjects are selected, the parameters are loaded again. This facilitates the reproduction of tables and graphs.

Table 6. Menus and adjustments for the window: "analysis"(presentations of results).

Level	The level of aggregation can be selected in this window (Switzerland) inventory unit, ...).
Estimator	In this area the type of estimator can be selected (target variable, reference unit, ratio estimators). The target variable and reference unit is shown as estimates of totals. For non-continuous data, the frequency of the occurring code values is calculated (e.g., as the number of stems or area in hectares).
Standard error	The standard error is available as an absolute value and has the same units as the variable value (timber cubic meter solid, hectare, number of stems, ...). When displayed as a percentage value the standard error is calculated in percentage of the variable value.
Presentation	In the area of *presentation*, it is possible to select as the output either table or graphic. The arrangement of the values within the table of graphic can be adjusted.
Output	Possible output media area: Printer, computer screen, and file.
Table arrangement	The window *table arrangement* allows selecting the arrangement of the elements in a table. The arrangement of the grouping element can be selected individually. Example for grouping elements: Reference unit Inventory identification number Target variable Assessment unit 1 Assessment unit 2 The arrangement of the graphs is similar.

The overall concept of the database, analysis, and presentation is presented in Figure 3.

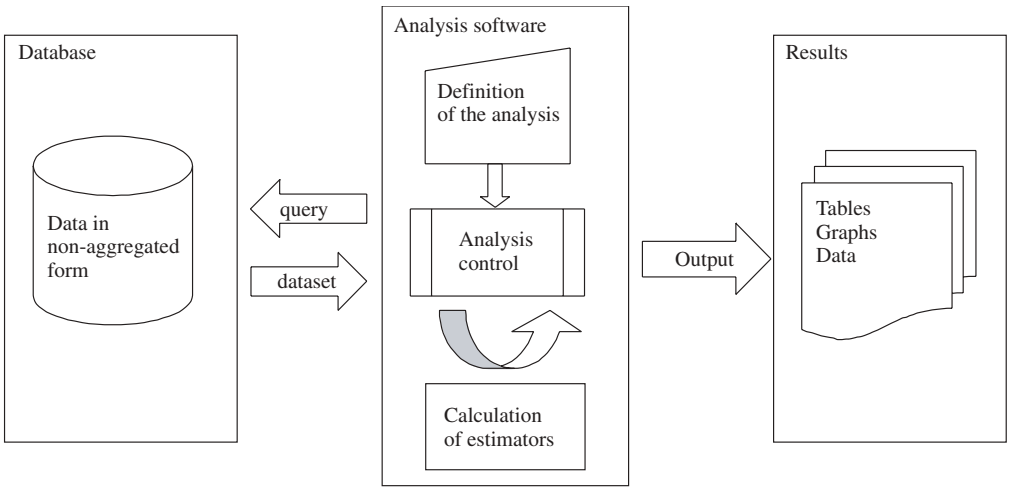

Figure 3. Analysis concept of the Swiss National Forest Inventory.

Other "Features" of the Analysis Software

Apart from programs for the file management within the SAS system and at the ORACLE or UNIX level, a data set can be generated according to the variables identified in the definition of the subjects. The results are: a) a comma delimited ASCII file and b) a temporary SAS file "WORK.AUSZUG." In order to work with these results in EXCEL the estimates and standard errors of the estimates can be exported. A semicolon delimited ASCII file is generated as an export file. For the structure of the data set and for the export file see Tables 7 and 8.

The batch mode makes it possible to conduct several analyses, one after the other. The batch can be run either with the user interface (online) or without the user interface (offline). The batch mode is started for the online mode directly from the analysis software. No other actions are possible until the end of the analysis. The systems messages appear in the LOG window.

Alternatively, it is possible to start the analysis without the window in the background. The system messages are then stored in a separate log-file. This file is overwritten each time a new analysis is run.

Table 7. Structure of the data set queried from the databank.

Variable	Definition
x	X-coordinate
y	Y-coordinate
clus	Identification number of the sample plot.
aussage	inventory unit
<befundeinh.1>	1. Assessment unit
<befundeinh.2>	2. Assessment unit
zielvariable	Coded target variable for categorical data; continuous value for the target quantity.
wicht	Weighting factor for the target variable.
begro	Reference unit

Table 8. Structure of the export files.

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

4.2.4 Literature

KÖHL, M. 1994: Statistisches Design für das zweite Schweizerische Landesforstinventar: Ein Folgeinventurkonzept unter Verwendung von Luftbildern und terrestrischen Aufnahmen. Mitt. Eidgenöss. Forsch.anst. Wald Schnee Landsch. 69 (1):141.

MATTISON, R. 1997. Data warehousing and data mining for telecommunications. Norwood, MA: Artech House, Inc. 273 p.

SAS INSTITUTE, INC. 1989. SAS/AF(R) Software: Usage and reference, Version 6. 1 ed. Cary, NC: SAS Institute, Inc. 245 p.

SAS INSTITUTE, INC. 1990. SAS(R) Screen Control Language: Reference, Version 6. 1 ed. Cary, NC: SAS Institute, Inc. 464 p.