Forestry Department
Food and Agriculture Organization of the United Nations

NATIONAL FOREST ASSESSMENT
CAMEROON

MANUAL FOR DATA PROCESSING AND ANALYSIS

DRAFT

Rome, February 2005

Forest Resources Assessment Programme **Working Paper 98**
Rome, 2005

The Forest Resources Assessment Programme

Forests are crucial for the well being of humanity. They provide foundations for life on earth through ecological functions, by regulating the climate and water resources and by serving as habitats for plants and animals. Forests also furnish a wide range of essential goods such as wood, food, fodder and medicines, in addition to opportunities for recreation, spiritual renewal and other services.

Today, forests are under pressure from increasing demands of land-based products and services, which frequently leads to the conversion or degradation of forests into unsustainable forms of land use. When forests are lost or severely degraded, their capacity to function as regulators of the environment is also lost, increasing flood and erosion hazards, reducing soil fertility and contributing to the loss of plant and animal life. As a result, the sustainable provision of goods and services from forests is jeopardized.

In response to the growing demand for reliable information on forest and tree resources at country and global levels, FAO initiated a programme to provide support to national forest assessments (NFA). The programme includes developing a harmonized approach to NFAs, information management and support to policy impact analysis for national level decision-making.

The purpose of the initiative is to introduce countries to an alternative approach designed to generate cost-effective information on forests and trees outside forests, including all benefits, uses and users of the resources and their management. Special attention is placed on monitoring the state and changes of forests, and on their social, economic and environmental functions. Another main objective is to build national capacities and harmonize methods, forest related definitions and classification systems among countries.

The Forest Resources Assessment Programme is organized under the Forest Resources Division (FOR) at FAO headquarters in Rome. Contact persons are:

Mohamed Saket, Forestry Officer, Mohamed.Saket@fao.org
Dan Altrell , Forestry Officer, Dan.Altrell@fao.org
Anne Branthomme, Forestry Officer, Anne.Branthomme@fao.org

or use the e-mail address: fra@fao.org

Compiled by John Fonweban Ngong in collaboration with Dan Altrell and Anne Branthomme.

DISCLAIMER

The Forest Resources Assessment (FRA) Working Paper Series is designed to reflect the activities and progress of the FRA Programme of FAO. Working Papers are not authoritative information sources – they *do not* reflect the official position of FAO and should not be used for official purposes. Please refer to the FAO forestry website (www.fao.org/forestry) for access to official information.

The FRA Working Paper Series provides an important forum for the rapid release of preliminary findings needed for validation and to facilitate the final development of official quality-controlled publications. Should users find any errors in the documents or have comments for improving their quality they should contact fra@fao.org.

Table of contents

List of figures

List of tables

Abbreviations

cc	canopy cover	**GPS**	Global Positioning System
Dbh	diameter at breast height	**LUS**	land use/forest type section
Dsh	diameter at stump height	**NGO**	Non Governmental Organization
FAO	Food and Agricultural Organization	**NWFP**	non wood forest product
FRA	Forest Resources Assessment programme	**P/S**	products and services
		RRA	rapid rural appraisal

1 Introduction

FAO has designed a sampling methodology based on a nation-wide low-cost field sampling, which while providing information on the present status of forest resources, also allows for long term monitoring of resources via a network of permanent sample plots. This design should be interesting, especially to developing countries whose knowledge base of their forest resources is still either completely lacking, incomplete or fragmentary and inappropriate for forest policy development; due lack of finances and/or technical expertise to undertake forest resource assessments.

FAO/FRA has as mandate to assist countries, at their request, in the provision of reliable information on forest resources and trees outside forest (TOF) base on this low-cost sampling approach.

In Cameroon, forest resource inventories have been undertaken mostly in the southern, predominantly forest region, while the northern, savanna region never been assessed. The Ministry in charge of forests in Cameroon saw the need for a nationwide resource assessment and requested the assistance of FAO to plan and undertake this assessment nation wide. An agreement was then signed in July 2002 with the following objectives:

- to plan and implement the first phase of a national forest assessment (NFA);
- to establish a monitoring system of forest and tree resources in Cameroon, and;
- to train the national team in NFA and in forest information management.

The objective of the assessment was to provide quantitative and qualitative information on the state of forest resources, their uses, management and change. Data are collected on a wide range of variables on forest and tree resources (biophysical characteristics of the resources and socio-economics aspects) to assess state of the forest as well as to get information on the use, users and management of the resources.

The method of data collection is through field measurements and interviews of local population and key informants.

Modifications and adaptations were made on the original methodology, to better serve Cameroonian national needs and taking into account the specificity of the countries ecosystem. Cameroon has two distinct ecological zones: the southern humid forest zone and the northern, predominantly dry savanna region, with gallery forests.

This presentation briefly outlines the sampling method put in place in Cameroon, database structure and data management, analysis and processing.

2 Characteristics of Sampling design used for the NFI in Cameroon

2.1 Description of the statistical design

The sampling design adopted for the National Forest Inventory (NFI) in Cameroon is systematic and stratified.

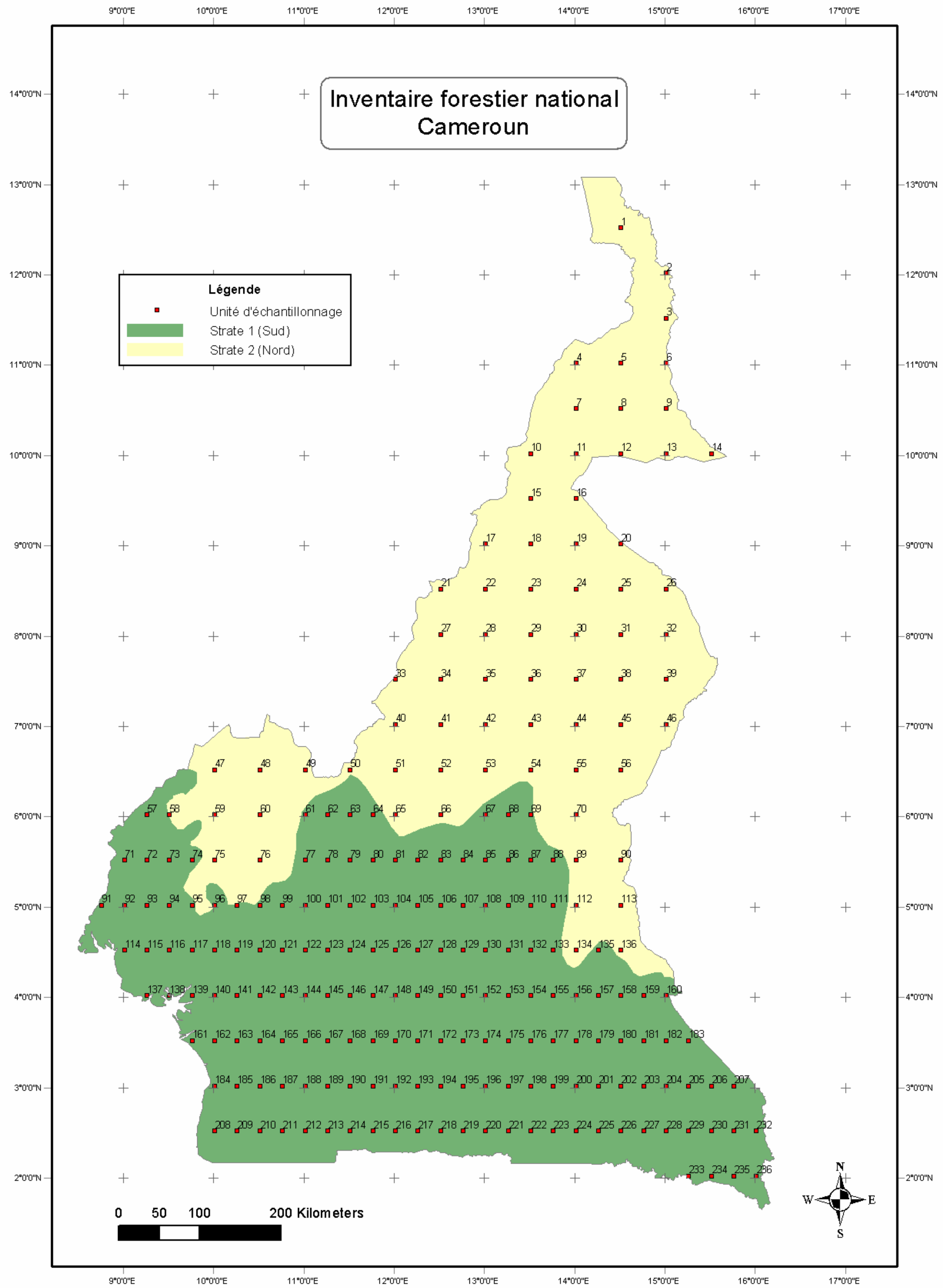
The country was first divided into two strata, according to ecological zones: a northern stratum, mainly consisting of savanna woodland and a southern stratum made of tropical forest. Two different sampling intensities were applied according to the strata. In the first stratum (South), a systematic grid of 30 minutes in latitude by 15 minutes in longitude was applied, while in the North, sampling units ("*Unités d'Echantillonnage*" or tracts) were selected at every 30 minutes latitude and 30 minutes longitude.

A total of 236 sampling units were envisaged for the National Forest Inventory (see Table 1 and Figure 1).

Table 1. Tract density per stratum in Cameroon

Stratum	**Tract number**	**Distance between tracts**	
		(minutes)	**(km)**
1	167	lat 30‘ x long 15‘	km 50 N X 25 km E
2	69	lat 30‘ x long 30‘	km 50 N X 50 km E
TOTAL	236		

Figure 1. Distribution of tracts for the national forest inventory of Cameroon



2.1.1 Tract description

The tracts or sampling units are square of 1km by 1km (100 ha). The co-ordinates of the south-west corner of the tracts correspond to those of the points selected in the systematic sampling frame.

At the centre of each sampling unit, a group of 4 rectangular plots or recording units ("*placettes*") of 20m by 50m (0.5 ha.) was established, a square of 500m x 500 m was defined. They start at each corner of an inner 500 m square (same centre as tract's), and are numbered clockwise from 1 to 4 as shown in figure 2. The location and orientation of the four plots are given in Figure and Table 2.

Figure 2: Tract with delineation of forest types and land use classes

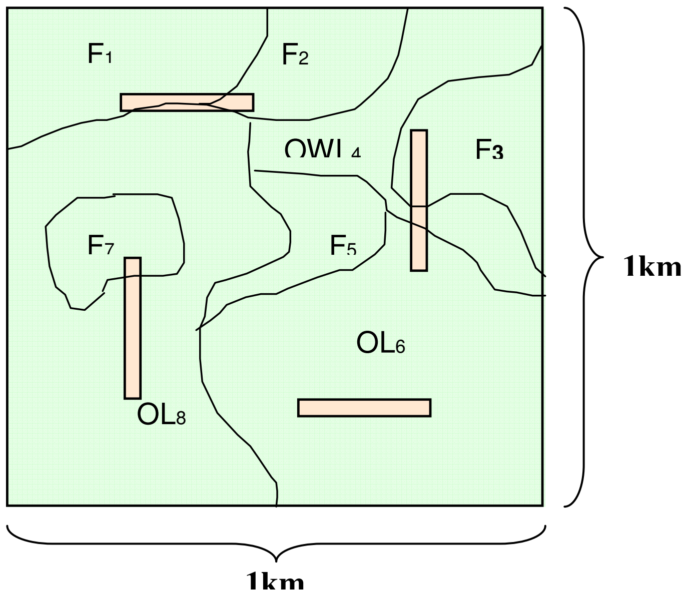
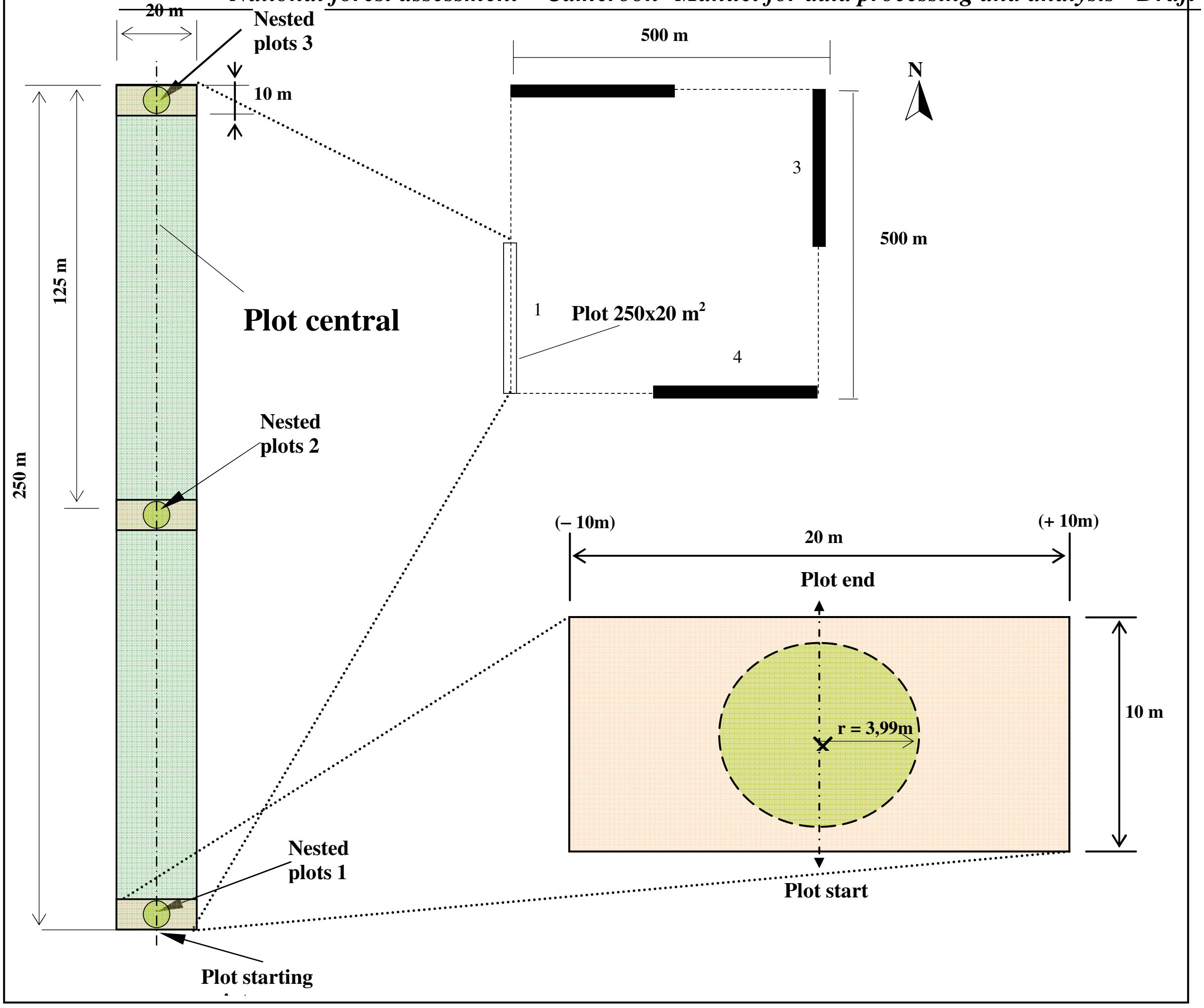


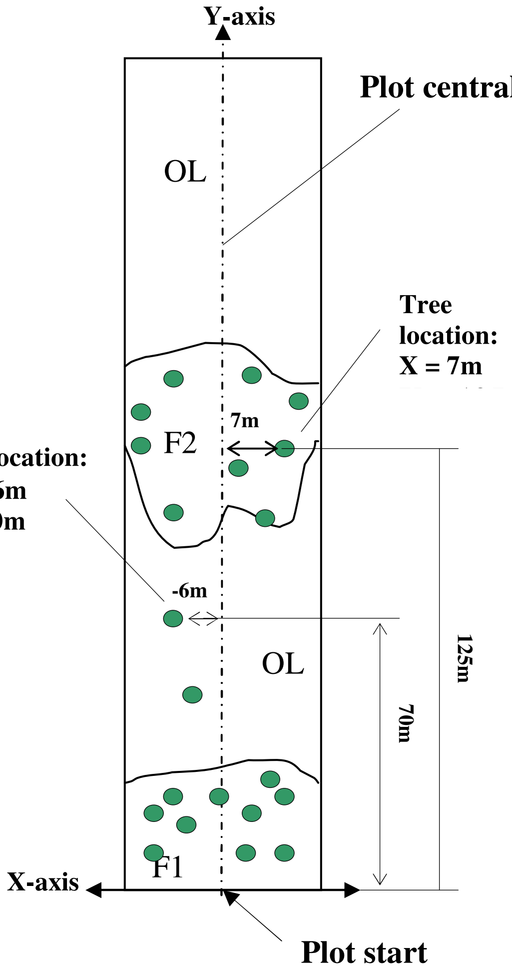
Table 2: Plot location and orientation

Plot	Location of the starting point of the plot, within the 500 m inner square	Orientation	Bearing
Plot 1	South-West corner	South-North	0 / 360 degrees
Plot 2	North-West corner	West-East	90 degrees
Plot 3	North-East corner	North-South	180 degrees
Plot 4	South-East corner	East-West	270 degrees

Furthermore, six small plots or sub-plots were established within each plot: three small rectangular plots, of 20m by 10m (200m²) and three circular plots of radius 3.99m (≈ 50m²) at the centre of the small rectangular plots.

During the survey, each plot was divided into land use sections (LUS) corresponding to homogenous forest/land use types.


20 m
Nested
plots 3
10 m
500 m
N
3
500 m
125 m
Plot central
1
Plot 250x20 m²
4
250 m
Nested
plots 2
(– 10m)
20 m
(+ 10m)
Plot end
10 m
r = 3,99m
Nested
plots 1
Plot start
Plot starting


Y-axis
Plot central
OL
Tree
location:
X = 7m
7m
F2
Tree location:
X = –6m
Y = 70m
-6m
OL
125m
70m
F1
X-axis
Plot start

2.1.2 Data collection

Details of data collection are found in the Manual of NFI for Cameroon (Branthomme 2003). Information was collected at the different levels:

- At tract level, data on population and access to the tract is recorded.
- At plot level, tree data (species, dbh, total height, commercial height, stem quality, and health status) was collected for trees with dbh ≥ 20 cm or dbh> 10cm for trees outside forest (TOF). Information on land use types was also collected, with estimates of sectional length and width for each land use type made.
- Information on the use of forest products and services, stand structure and management, as well as general information concerning the area were collected for each land use type; mostly through interviews. Observations on wildlife indices (direct observations, vocalizations, traces, etc...) were also noted at sampling unit level.
- At sub-plot level, information on soil types, topography and regeneration was collected. Regeneration was defined as trees with dbh<10cm found within the circular plots, while saplings ("*gaulis*") or trees with diameter between 10 and 20cm (10≤ dbh< 20cm) were counted within each rectangular sub-plot. All trees were botanically identified as much as possible to species level by tree spotters/botanists (*"prospecteurs"*).

2.2 Considerations into the choice of Sampling Design

2.2.1 Rationales behind the sampling approach

The sampling design was built up around the following principles (Saket et al, 2002):

- Stratification based on forest types and land use classes is discarded as strata may change over time and render stratified sampling irrelevant with permanent samples. However, stratification on basis of relatively stable strata such as ecological zoning may be applied to countries with pronounced differences in ecosystems, such as humid and dry forests (Cameroon).
- The sampling is designed to include tree resources outside forest and an array of biophysical and management/uses parameters of forest and trees outside forest.
- The approach foresees establishing permanent sample plots to monitor changes over time between forests and other land use classes and the processes within the forests;
- Biophysical data and the management and use information is geo-referenced and integrated to enable assessing the size and state of the resources in relation to their geographic location in the country and their social context;
- Data quality is guaranteed and measurement methods are harmonized.

2.2.2 Advantage of a systematic sampling design compared to a random sampling design

For a number of motives including practicality, data quality, cost implications, the systematic sampling was chosen for forest and tree data collection (Saket et al, 2002). The systematic sampling in forest inventory is: (i) easily planned, (ii) faster in execution and mostly cheaper; (iii) it gives better estimates of the mean than unrestricted random sampling and even stratified random sampling in large areas, because the variation which may be considerable in such areas is better represented in the sample for which the distribution of the sample plots within the surveyed population is homogeneous; (iv) it gives thus better precision compared to random sampling.

2.2.3 Cluster layout

The cluster or multi-stage sampling is statistically sound and effective design, frequently used in forest inventories. It is less costly than simple or stratified random sampling if the cost of obtaining observations increases as the distance separating the elements increases (Schaeffer et *al*. 1990). This is true in forest sampling, where at times location and getting to a sampling unit is expensive, while measurement of the unit is relatively cheap (Freese, 1962). This advantage is particularly evident in areas of difficult accessibility such as the tropical primary forests. Another advantage is the reduction of systematic errors or bias due to efficient control of field operations. A major inconvenience resides in the lost of precision compared to that obtained from the simple random sampling (Rondeux, 1993).

In order to minimise the effects from systematic or periodic variations within the surveyed populations, the 4 plots are placed in perpendicular orientations. This is also proven practical in the fields as the field crews, during the survey, progress from the first plot to the last one without walk backwards for measurements. The way back to the road is easier to find.

2.2.4 Plot structure

The plot was designed to cross the maximum possible of variations within and between the classes, but long strips are discarded for practical reasons.

2.2.5 Sampling intensity

A total of 236, 1Km x 1Km square sampling units were envisaged for the NFA in Cameroon. However, in each sampling unit only 2 ha (four sub-plots of 0.5 ha), giving a total sampled area of 472 ha (2 ha x 236).

The total country area is 47,544,000 ha, the stratum 1 (South) covers 26,346,593 ha, (55% of the country) while the stratum 2 (North) amounts 21,197,407 ha (45%). This gives a sampling intensity of 0.0013% for the stratum 1, 0.0007% for the stratum 2 and 0.0010% for the whole country. This is a very low sampling intensity.

This is however deemed to yield estimates of forest and land use attributes at acceptable precision at the national level. Precision of less abundant attributes decreases with the decrease of their frequency of occurrence in the country.

3 Statistical methods for the National Forest Assessment

3.1 Type of statistical design

The preceding description reveals some degree of stratification, a systematic arrangement of sampling units, and some hierarchical nature in the sampling design; in which each level or stage of sampling is nested within another (from sampling unit/tract to plots and sub-plots). In forestry, one would think of sub-sampling (Cochran 1977, Freese 1962) or multi-stage sampling. In experimentation, one would attribute this to a nested or hierarchical design (Bishop; Box et *al*. 1978).

The sampling design (plan) could most probably be described as a stratified systematic two-stage sampling design, but could also be thought of as a stratified two-stage cluster sampling design (Scheaffer et *al*. 1990). Although some forestry/sampling literature has tried to distinguish between two-stage sampling and cluster sampling, these terms have always been used interchangeably, referring almost to the same design. The following examples will help to illustrate this:

- Loetsch and Haller (1973: p139-163) talks of the cluster sampling in terms of layout (arrangement) of sampling or recording units ("clustered around camps"), but further goes to describe a two-stage sampling design as follows:

"**Cluster sampling** implies the division of the population into units of hierarchic order. Each step represents a sampling stage. The design may include two, three or more stages. In **a two-stage sampling** primary sampling units will be selected from the population of *M* primary units in the first stage. From each of *M* selected primary units *m* sampling units will be selected from the population of *N* secondary units in the second stage..."

- Rondeux (1993) talks of multi-stage sampling and considers cluster sampling ("échantillonnage par grappes") as a special case of multi-stage sampling.
- Dussaix and Gros-becs (1993) also describes "sondage à deux degrés/par grappe" interchangeably.
- Husch et al (1982), after describing multi-stage/two– stage sampling states that another frequently used two-stage sampling design in forest inventory employs groups or clusters of plots or sampling points at random location.
- Scheafer et al (1990, Chap. 9; p.285-310), explicitly uses the term two-stage cluster sampling, and consecrates an entire chapter for it. As they defines it:

"A two-stage cluster sample is obtained by first selecting a probability sample of clusters and then selecting a probability sample of elements from each sampled Cluster".

- Freese (1962: p50-60) describes two-stage sampling as used in forestry. However, he makes mention of the fact that where secondary units are "arranged in a set pattern" or displayed in a systematic fashion, it is no longer two-stage sampling in the sense of the term; but might be called cluster sampling, the cluster being the group of secondary units at each location.
- Avery and Burkhart (1994), citing Scheaffer et al. (1990) talks of cluster sampling, but ends up with formulae for two-stage sampling from Freese (1962).

The above examples from literature indicate how close the two designs may be; and the confusion this can bring when it comes to choosing an appropriate sampling design. In the case of the National Forest Inventory in Cameroon one would be inclined to think that the a stratified two-stage cluster sampling design (or cluster sampling design??) was used for the following reasons: the arrangement of the secondary units in a regular pattern (within the 500m x 500m square) can be likened to a cluster. Furthermore, the description in the field manual indicates that the 4.plots are "*a cluster*" or a group (=cluster) of recording (observation) units. Even though, cluster and two-stage sampling are not synonymous, they can co-exist because it is possible to conceive a two-stage sampling design in which each stage consists of clusters. This is probably the case with the present NFI in Cameroon. However, in order to decide on which formula to use for statistical calculations, it is necessary to examine the statistical formulae proposed by the various authors for two-stage and two-stage cluster sampling.

3.2 Formulae for two-stage and two-stage cluster sampling

The **mean** for a two-stage sampling is *given* by the expression:

$$\bar{\bar{y}} = \frac{\sum_{i=1}^{n} \bar{y}_i}{n} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m} y_{ij}}{mn}$$

Various expressions have been used to define the **variance** for a two-stage sampling or two-stage cluster sampling or cluster sampling survey by different authors.

An analytical comparison of the different mathematical formula used show, that the formulae proposed Loetsch et al (1973), Rondeux (1993), Cochran (1977), Freese (1962), Avery and Burkhart (1993), and Scheaffer et *al*. (1990) are the same, expressed in different way (see Annex).

A computational comparison was made using the formulae and the results are presented in table…. The results give the same standard errors using the various formulae. We may thus conclude that two-stage sampling and two-stage cluster sampling are different terminology to describe the same sampling.

Table 1. Computational comparison of various two-stage/cluster sampling formulae

Loetsch & Haller (1973); Rondeux (1994)		FREESE (1962)		SCHEAFFER et al. (1990)	
Volume		Volume		Volume	
Between Variation	11303.948	**Between Variation**	48986.0721	**Between Variation**	48986.0721
Within Variation	3770.281	**Within Variation**	3770.281	**Within Variation**	3770.281
Varcomp1	313.9747578	**Varcomp1**	48982.35947	**Varcomp1**	340.1552741
Varcomp2	26.18246726	**Varcomp2**	0.28003266	**Varcomp2**	0.001944671
Total Variance	340.157225	**Total Variance**	340.1572188	**Total Variance**	340.1572188
Standard Error	18.443352	**Standard Error**	18.44335161	**Standard Error**	18.4433516
Mean	213.844	**Mean**	213.444	**Mean**	213.844

Notes: *Freese (1962) talks of Two Stage Sampling; Loetsch talks of cluster sampling and two-phase sampling; Scheaffer et al talk of Two-Stage Cluster Sampling, and the formula is for equal size clusters (number of elements and selected elements in the cluster)*

The formula for the variance proposed Loetsch et al. (1973) and Rondeux (1993) will be presented as reference:

$$S_y^2 = (1 - \frac{m}{M}).\frac{s_\beta^2}{m} + (1 - \frac{mn}{MN}).\frac{s_\gamma^2}{mn}$$

where:

s_β^2 =variability (variance) between primary units;

s_γ^2 =variability (variance) between secondary units within primary units;

M = Total number of primary units (Pus) in the population;

m = Number of primary units contained in the sample (number of Pus sampled);

N = Total number of secondary units in each primary unit or within each primary; and

n = number of secondary units sampled in each primary unit selected.

$s_{ß}^2$ and s_{γ}^2 are obtained from the following formulae:

$$s_\gamma^2 = \frac{\sum_{j=1}\sum_{i=1} y_{ij}^2 - \frac{\sum_j (\sum_i y_{ij})^2}{n}}{m(n-1)} \quad \text{and} \quad s_\gamma^2 + ns_\beta^2 = \frac{\frac{\sum_j (\sum_i y_{ij})^2}{n} - \frac{(\sum_j \sum_i y_{ij})^2}{mn}}{m-1}$$

In the case of the NFA, the primary units are the tracts, while the secondary units are the plots (or subplots).

Ratio Estimator for Stratified Sampling:

If the number of n_h sampling units in the strata are large enough and if the ratios R_h in the various strata are different enough, it is demonstrated that the best estimates of the raio and of the variance of the estimates are as follows:

- Estimate of the mean per size unit (“ separate ratio estimate”)

$$\overline{R}_{1s} = \sum_{h=1}^{L} \frac{X_h}{X} \overline{R}_{1h} = \sum_{h=1}^{L} \frac{X_h}{X} \frac{\overline{y}_h}{X_h}$$

- Estimate of the variance of R1s :

$$v(\overline{R}_{1s}) = \frac{1}{X^2} \sum_{h=1}^{L} \frac{N_h}{N^2} \frac{1-f_h}{n_h(n_h-1)} (\sum_{i=1}^{n_h} y_{hi}^2 + \overline{R}_{1h}^2 \sum_{i=1}^{n_h} x_{hi}^2 - 2\overline{R}_{1h} \sum_{i=1}^{n_h} x_{hi} y_{hi}$$

Two stage sampling with secondary units of unequal sizes and no stratification (ratio estimates)

Ratio estimation can be used here like in one-stage sampling designs, the size of a secondary unit being the auxiliary parameter.

- Estimate of mean value per size unit:

$$\overline{R}_2 = \frac{\sum_{i=1}^{n} \overline{y}_i}{\sum_{i=1}^{n} \overline{x}_i} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m} y_{ij}}{\sum_{i=1}^{n}\sum_{j=1}^{m} x_{ij}}$$

- Estimate of the variance of R2

$$v(\overline{R}_2) = \frac{1}{\overline{\overline{X}}^2} \frac{1-f_1}{n(n-1)} [\sum_{i=1}^{n} \overline{y}_i^2 + \overline{R}_2^2 \sum_{i=1}^{n} \overline{x}_i^2 - 2\overline{R}_2 \sum_{i=1}^{n} \overline{x}_i \overline{y}_i] +$$

$$\frac{1}{\overline{\overline{X}}^2} \frac{f_1(1-f_2)}{nm} \sum_{i=1}^{n}\sum_{j=1}^{m} \frac{[(y_{ij}-\overline{y}_i)^2 + \overline{R}_2^2 (x_{ij}-\overline{x}_i)^2 - 2\overline{R}_2 (y_{ij}-\overline{y}_i)(x_{ij}-\overline{x}_j)]}{n(m-1)}$$

3.3 Statistical formula for simple sampling

In the National Forest Assessment, simple random statistical formulae were used to simplify calculations. The tract is the sampling unit, so all information from plot is grouped at tract level.

The estimators to be used in the NFA are in general ratio estimators so the statistical expressions are given for ratio estimates in Frame 1.

Frame 1. Statistical expressions for ratio estimates for simple sampling

Ratio estimate (R) :

$$\mathbf{R} = \frac{\bar{x}}{\bar{y}} = \frac{\frac{1}{n}\sum_1^n x_i}{\frac{1}{n}\sum_1^n y_i} = \frac{\sum_1^n x_i}{\sum_1^n y_i} \Rightarrow (\bar{x}) = R * \bar{y} = \frac{\sum_1^n x_i}{\sum_1^n y_i} * \bar{y} \Rightarrow (X) = Y * R = Y\frac{\sum_1^n x_i}{\sum_1^n y_i}$$

Where $\bar{x}$ is the mean of the measured variable (ex. tree volume in forest area) in the sample and $\bar{y}$ is the mean of the auxiliary variable (ex. forest area) in the sample.

Variance of population (V) :

$$\mathbf{V} = \frac{\sum_1^n (\mathrm{x} - \bar{x})^2}{n-1} = \frac{\sum_1^n (\mathrm{x} - \mathrm{R} * y)^2}{n-1} = \frac{\sum_1^n \left(\mathrm{x} - \frac{\sum_1^n x_i}{\sum_1^n y_i} * y\right)^2}{n-1} = \frac{\sum_1^n (x_i)^2 + R^2 \sum_1^n (y_i)^2 - 2R\sum_1^n x_i\, y_i}{n-1}$$

Where n is the sample size for the estimation of the variable (ex. to estimate tree volume in forest area, n is the number of sample units with forest area)

Standard deviation of population (Sd):

$$\mathbf{Sd} = \sqrt{V} = \sqrt{\frac{\sum_1^n (\mathrm{x} - \mathrm{R} * y)^2}{n-1}} = \sqrt{\frac{\sum_1^n \left(\mathrm{x} - \frac{\sum_1^n x_i}{\sum_1^n y_i} * y\right)^2}{n-1}} = \sqrt{\frac{\sum_1^n (x_i)^2 + R^2 \sum_1^n (y_i)^2 - 2R\sum_1^n x_i\, y_i}{n-1}}$$

Coefficient of Variation (CV%):

$$\mathbf{CV\%} = 100\frac{Sd}{\bar{x}} = 100\frac{\sqrt{V}}{\bar{x}} = 100\frac{\sqrt{\sum_1^n (\mathrm{x} - \mathrm{R} * y)^2}}{\bar{x}\sqrt{(n-1)}} = 100\frac{\sqrt{\sum_1^n \left(\mathrm{x} - \frac{\sum_1^n x_i}{\sum_1^n y_i} * y\right)^2}}{\bar{x}\sqrt{(n-1)}} = 100\frac{\sqrt{\sum_1^n (x_i)^2 + R^2 \sum_1^n (y_i)^2 - 2R\sum_1^n x_i\, y_i}}{\bar{x}\sqrt{(n-1)}}$$

Variance of Ratio estimate ($\mathbf{V_R}$):

$$\mathbf{V_R} = \frac{1}{(\bar{y})^2} * \frac{V}{n} = \frac{\sum_1^n (\mathrm{x} - \mathrm{R} * y)^2}{\left(\frac{\sum y_i}{n}\right)^2 * n(n-1)} = n * \frac{\sum_1^n \left(\mathrm{x} - \frac{\sum_1^n x_i}{\sum_1^n y_i} * y\right)^2}{\left(\sum y_i\right)^2 * (n-1)} = n * \frac{\sum_1^n (x_i)^2 + R^2 \sum_1^n (y_i)^2 - 2R\sum_1^n x_i\, y_i}{\left(\sum y_i\right)^2 * (n-1)}$$

Standard error of Ratio estimate ($\mathbf{S_R}$):

S

S_R

$$= \sqrt{V_R} = \sqrt{\frac{\sum_1^n (\mathrm{x} - \mathrm{R} * y)^2}{\left(\frac{\sum y_i}{n}\right)^2 * n(n-1)}} = \sqrt{n * \frac{\sum_1^n \left(\mathrm{x} - \frac{\sum_1^n x_i}{\sum_1^n y_i} * y\right)^2}{\left(\sum y_i\right)^2 * (n-1)}} = \sqrt{n * \frac{\sum_1^n (x_i)^2 + R^2 \sum_1^n (y_i)^2 - 2R \sum_1^n x_i y_i}{\left(\sum y_i\right)^2 * (n-1)}}$$

Sampling error (SE_R):

SE_R

$$= t * S_R = t\sqrt{\frac{\sum_1^n (\mathrm{x} - \mathrm{R} * y)^2}{\left(\frac{\sum y_i}{n}\right)^2 * n(n-1)}} = t\sqrt{n * \frac{\sum_1^n \left(\mathrm{x} - \frac{\sum_1^n x_i}{\sum_1^n y_i} * y\right)^2}{\left(\sum y_i\right)^2 * (n-1)}} = t\sqrt{n * \frac{\sum_1^n (x_i)^2 + R^2 \sum_1^n (y_i)^2 - 2R \sum_1^n x_i y_i}{\left(\sum y_i\right)^2 * (n-1)}}$$

Rel. Sampling error (SE_R%):

SE_R%

$$= 100t\frac{S_R}{\bar{x}} = 100t\frac{\sqrt{\sum_1^n (\mathrm{x} - \mathrm{R} * y)^2}}{\bar{x} * \sqrt{\left(\frac{\sum y_i}{n}\right)^2 * n(n-1)}} = 100t\frac{\sqrt{n * \sum_1^n \left(\mathrm{x} - \frac{\sum_1^n x_i}{\sum_1^n y_i} * y\right)^2}}{\bar{x} * \sqrt{\left(\sum y_i\right)^2 * (n-1)}} = 100t\frac{\sqrt{n * \sum_1^n (x_i)^2 + R^2 \sum_1^n (y_i)^2 - 2R \sum_1^n x_i y_i}}{\bar{x} * \sqrt{\left(\sum y_i\right)^2 * (n-1)}}$$

<u>STRATIFIED SAMPLING</u>:

$$Total\ Population(N) = \left(\frac{Total\ area}{Plot\ size}\right) = N_1 + N_2 + ... + N_k$$ (Where k = *number of strata)*

Sample of population, sample size: $(n) = n_1 + n_2 + ... + n_k$

Estimate of overall population mean: $$(\bar{X}) = \frac{\sum_{t=1}^{k} N_t \bar{x}_t}{N} \quad \left(Where\ \bar{x}_t = \sum_{j=1}^{n_t} \frac{x_{tj}}{n_t}\right)$$

$$\left(W_{t\bar{x}}\right) = \frac{N_t}{N} \quad \left(= \frac{Area_{stratum_t}}{Area_{total}}\ in\ the\ case\ of\ LandUse\ area\ estimation\right)$$

4 National Forest Resources Assessment database

To store and manage the collected NFA data, a database application was developed by FAO/FRA in collaboration with the country national team.

The database application, based on the MS Access software (2000/2002), comprises two database files one "data" database (NFI-data_<*COUNTRY*>.mdb) and one "application" database (NFI-<*COUNTRY*> v.*x.x*.mdb). All collected field data (dynamic data) are stored in inter-related tables in the "data" database, while the "application" database contains code tables (static data), forms, queries and macros.

The users of the NFA database manage the data through the application database. To open the database application the users double-click on the *NFI-<COUNTRY> v.x.x.mdb* file. The data are managed through forms and the user navigates in the database by selecting different forms according to required operation.

The efforts in developing the NFA database application have initially focused on wise data storage and on facilitating the input of primary data, why until now the only activated section in the database application is the section on "*Field Data*".

The design of the forms for adding/editing field data in the database application follows the same structure than for the field forms employed for the data collection in the field inventory. This consistency facilitates data entry into the database application. Moreover, the instructions on how to register field data in the field forms, as given in the field manual, are also valid for recording the information into the database (see **Annex 1**).

4.1 Tables

All NFA data are stored in tables. Tables with inventory data (primary data) are kept in the "data" database (NFI-data_<*COUNTRY*>.mdb), while all static data like codes, expansion factors, etc. are stored in the "application" database (NFI-<*COUNTRY*> v.*x.x*.mdb) (refer to Figure 5).

4.1.1 Code tables

For each variable with attribute options, there is a corresponding code table with defined options. The names of code tables all begin with "C-" (e.g. "C-Fire", see **Figure 3**). These tables contain internationally harmonized and national terminology and options for the variables, as well as definitions, where every option has a unique code.

Figure 3. Example of Code table giving codes, options and definition for the Fire variable

C-Fire : Table

ID	Fire_ID	Fire_text_eng	Fire_text_spa	Fire_text_fre	Fire_	Fire_text_N	Fire	Fir	Fir	Fire_def_fre
1	1	No evidence	Sin evidencia	Pas de trace d'incendie	1	Pas de trace d'ince	1			Pas de trace d'incendie dans la SUT
2	2	Recent	Reciente	Incendie récent	2	Incendie récent	2			Traces d'incendie datant de la saison/année a
3	3	Old	Antiguo	Incendie ancien	3	Incendie ancien	3			Traces d'incendie datant de la saison/année p
ber)										

Record: 1 of 3

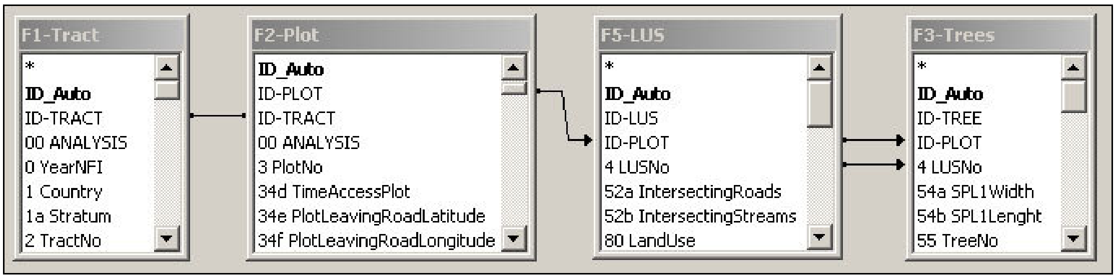
4.1.2 Data tables

For each level of data collection (Tract, Plot, Land Use Section, etc.) there is a corresponding data table with defined variables (or field). The variables are internationally harmonized and adapted to national terminology. They may have unbound values or bound to predefined attribute options according to the code tables. The names of data tables begin with "F-", "P-", "Ph-" or "H-"(see Figure 6). The *F-tables* contain the values of all the variables that are collected at corresponding inventory level: Tract data is stored in the *F1-Tract* table, Plot data is stored in the *F1-Plot*, LUS data is stored in the *F5-LUS* table, etc. The *P-table* contains data on informants or other persons involved in the inventory and the *H-tables* are help-tables that relate the data from two data tables to each other, for example the *H-Person-Function-Tract* relates the information of persons to tracts and indicates what function the person had in that tract.

4.2 Relation database

The NFA database application is developed in MS Access, which is a relation database. A relation database allows that data can be collected at different inventory levels (Tract, Plot, LUS, Tree, etc.) and still relate to each other. In practice, this is accomplished by creating separate data tables for each inventory level. To relate (link) the data in one table to another the tables must have at least one common field. In this way the tract attributes can be linked to the attributes in all plots in the tract through the "ID-TRACT" field, and the Plot attributes can be related to the attributes in all LUS in the plot through the "ID-PLOT" field, and the LUS attributes can be related to all product/service attributes in the LUS through the ID-LUS field and further to the attributes of all trees through the two common fields ID-PLOT and ID-LUS, etc (see Figure 4).

Figure 4: Illustration of the relationship between different data tables



Tables can be linked to each other even though they are not physically in the same database file. In consequence, a fixed location of the database files is necessary. As the database application is linked to the dynamic tables in the "data" database, it therefore needs to know where the "data" database is located. The NFA database application consists of the two database files "*NFI-data_<COUNTRY>.mdb*" and "*NFI-<COUNTRY> v.x.x.mdb*" and the location of the data database is set to the "C:\NFA\" folder (see Figure 5).

Figure 5: Location of database files

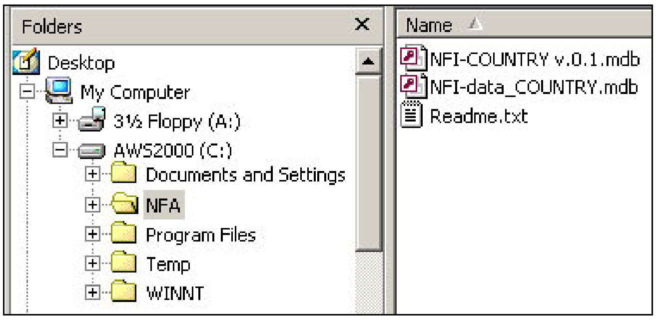
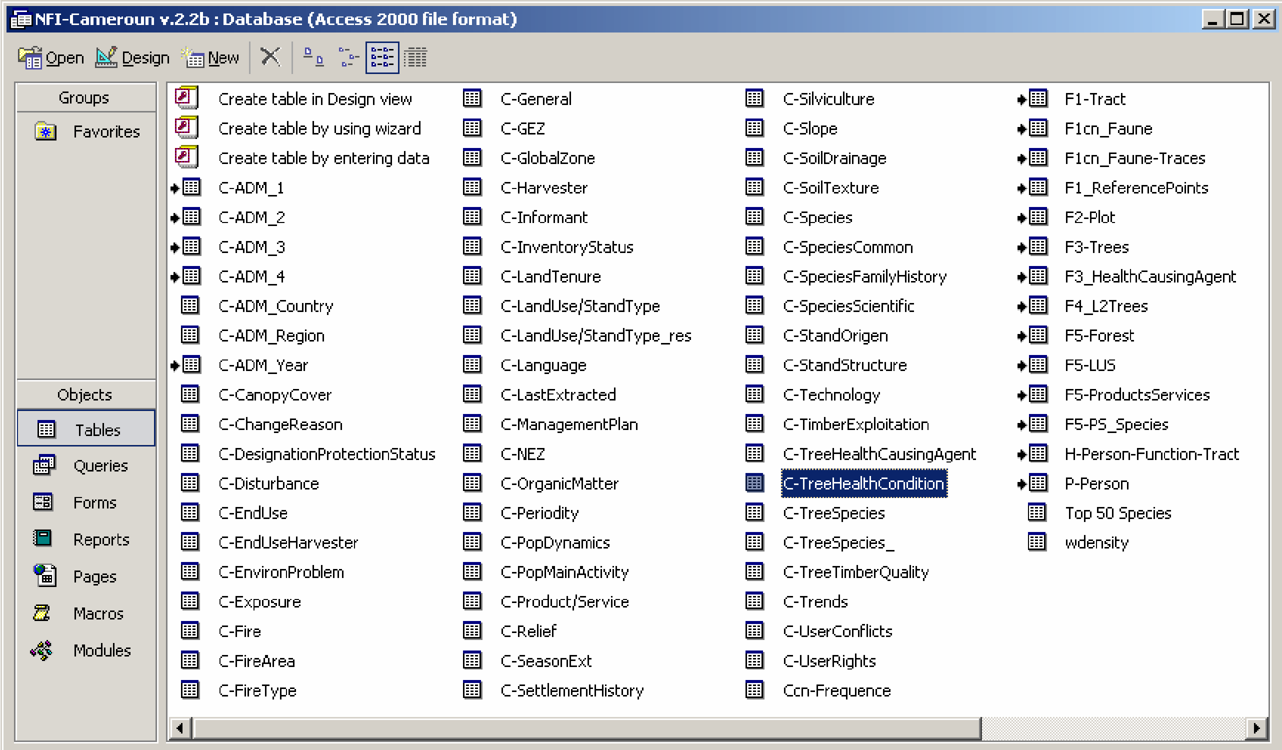


Figure 6 shows the database window from the application database (*NFI-<COUNTRY> v.x.x.mdb*) and displays some of the tables in the database application. The tables displayed with only a table symbol next to the table name are tables stored in the same application database, while the tables with an arrow symbol next to the table symbol are links to the tables with dynamic data (primary field data) that are stored in the data database (*NFI-data_<COUNTRY>.mdb*).

Figure 6: Database window in the NFA application database ("*NFI-<COUNTRY> v.x.x.mdb*")



Notes: The figure shows some of the code tables in the database (beginning with "C-") and some of the linked data tables (with an arrow symbol next to the table symbol).

The purpose of separating the dynamic field data from the rest of the database application is to allow users to work with the database application from different computers in a network and have only one main database with the inventory data.

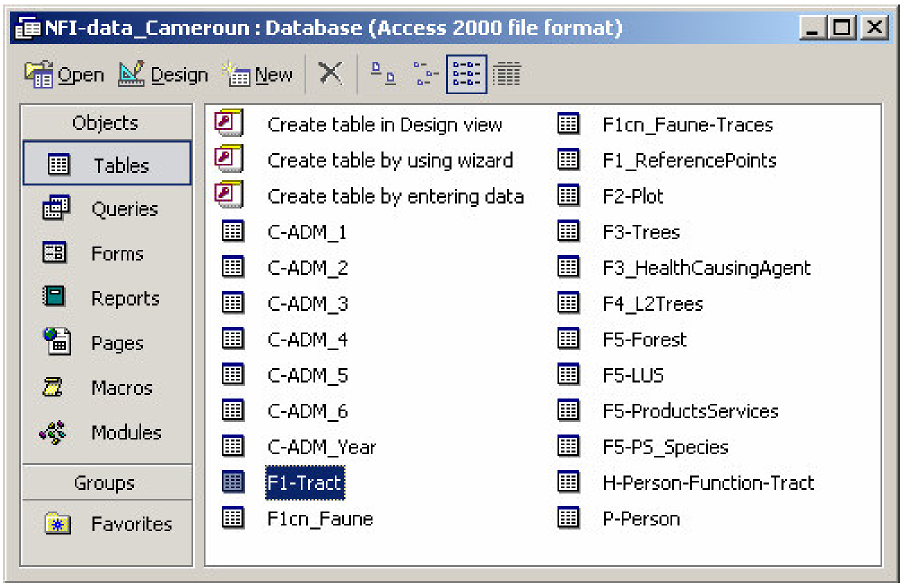
The NFA application database can be installed on every computer in the network and be related to a single data database where the inventory data are stored. In this way only one version of the inventory data exists. Another benefit from separating the dynamic data from the rest of the database is to facilitate the updates of features in the database application without altering the primary data. In this way, a database specialist can develop new versions of the NFA database application, with improved functionalities, while other users continue to enter/edit data through an older version of the database application. When the new version of the application database ("*NFI-<COUNTRY> v.x.x.mdb*") is ready it will substitute the older version, but the primary data remains unaltered in the data database (*NFI-data_<COUNTRY>.mdb*).

4.3 Forms

The users of the NFA database manage and enter the data through forms. There is one form corresponding to every field form used for field data collection. The form may be subdivided into sub-forms (see **Figure 7**). The users can add and edit the NFA field data through these forms. Validity domain of the field data is built-in in the forms. Criteria for acceptable values and reasonable values are predefined for most of the variables, so when the users enter field data through the forms they automatically go through a general validation so unrealistic field values are rejected.

More information on the forms and on how to navigate in the NFA database application is given in **Annex 1**.

Figure 7: Data Database window in the NFA application database ("NFI-<COUNTRY> v.x.x.mdb") showing some of the forms that are employed in the database application



4.4 Queries

Queries can be used for processing and analyzing data that are stored in tables (see following section). However, it is not recommended to process or analyze the NFA data in any of the two database application files, *NFI-data_<COUNTRY>.mdb* and *NFI-<COUNTRY> v.x.x.mdb*, as it would cause the application to become very "heavy". Instead, the data processing and analysis should preferably be carried out in a separate database that is linked to the two database application files.

In the NFA, database application the queries are only employed to filter the data in tables or in other queries. The information displayed through forms can be filtered through queries or through tables with applied filters. For each form, there can be one or more queries employed depending on the structure and functionality of the form. For example, the data source for every sub-form in a form could be based on a separate query.

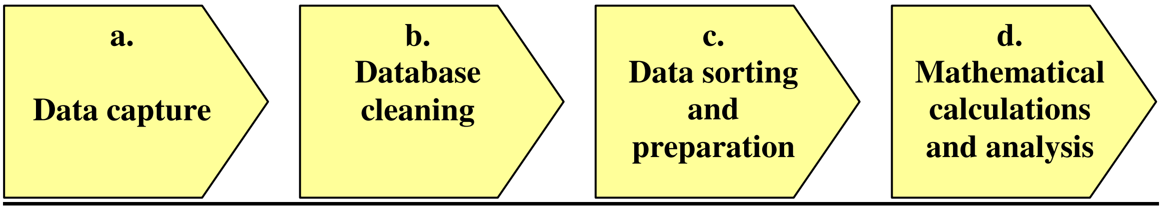
5 Data processing and analysis

Inventory data from any source must first be recorded, edited and condensed before it can be processed. In addition, data processing procedure must be adapted to the specific requirements and design of the inventory itself. It is therefore important that the treatment of data be considered as an integral part of a forest inventory from the very beginning planning stage. Data recording and processing should be given careful consideration at the initial planning stage of forest inventory, since the means of data processing or the cost may have a considerable impact on the design, intensity, and timing of the entire inventory.

Although data processing serves more as a tool of forest inventory than as a determining factor, its influence on the realization of the inventory should not be underestimated.

The **Figure 8** presents the different steps to conduct data processing and analysis.

Figure 8: Steps for data processing and analysis



5.1 Data capture

Data capture (or entry) consists in transferring field data, from paper field forms into digital format. Data entry for the NFA in Cameroon is done through the database application developed by the FAO and described in the previous section.

In order to facilitate the data entry work, the application interface is structured in a similar way than the field forms. Users can easily navigate between the different forms to enter or view the data collected in the field. Details on the use of the application are given in **Annex 1**.

Persons responsible for data entry need to be adequately trained on how to use the database. Regular backup of the data databases (tables with primary data) should also be made.

5.2 Data validation and cleaning

In the processing of any inventory data comparison and checking are essential at each phase to ensure integrity and reliability of information. Checks should be made to verify that all data were properly recorded in the field forms and then in the database. Furthermore, plausibility of the magnitude of figures must be checked.

Some sources of errors that need to be checked include:

- errors made during data collection (measurement errors, misinterpretation of variables and options, inconsistencies between field teams);
- errors made during data coding (transcription of the measurements and observations if the field forms, spelling mistakes of species name);
- errors introduced due to wrong reading of field forms during the data entry process;
- errors caused by pressing wrong keys during the data entry process;
- errors due to wrong programming etc.

While some errors may be minimized if careful checks are made, others can pass unnoticed and may lead to bias. It is often good to carry out checks with field team leaders (recorders and tree spotters) to reconcile certain species names that may have been wrongly written in the field. It is also possible to detect that there is an error if a tree 30 m tall has a diameter of 15 cm (most probably 150 cm); but it may not be possible if a tree with diameter 45 cm was recorded as 25 cm.

The checking of data results in i) checking the field forms, before the data entry work; ii) checking the data entry by controlling a given number of randomly selected records; iii) checking the tables containing primary data in the database, when it is ready for data analysis, using criteria, filters, routine and queries.

Checks result in the identification of errors of various sorts and inconsistencies and the necessary corrections are made in the database, to produce clean reliable data tables**.**

The data table serves as input file for all generation operations and calculation. Therefore, the cleaning process is essential to ensure that the data processing can be done smoothly and give expected results.

5.3 Data sorting and preparation for calculation

This step of the inventory data processing system includes all operations aiming at preparing the basic data for further computations. Sorting arranges data in a particular or specific manner for data processing. In the processing of large amounts of inventory data, this operation is important, as data has to be grouped using various criteria. The data must be sorted depending on the inventory design, on the type of data gathered in the field and on the various classifications scheme used (strata, inventory units...).

The storage of the primary data in a structured database facilitates this operation, since tables can easily be linked to each other. Database queries or routines are also powerful tools to compute data. One major advantage is that, once created, queries/routines can be applied on updated datasets to produce new results without additional efforts.

In the NFA in Cameroon, queries were prepared for each key variable to be assessed to generate a table than can easily be processed in Excel or any Statistical software to generate means and other statistical estimates.

5.4 Description of data sorting for each variable

5.4.1 Steps for estimate calculation, analysis and validation

This step consists in computing intermediate and final results from prepared, sorted data. Processing of inventory data involves mostly computations as well as statistical analysis (calculations of standard errors, including analysis of variance or regression equations...).

At any level of the study, the following estimates have to be computed and presented in multi-entry tables, related to land use/forest type classes, ecological zones, species or groups of species and to diameter classes:

- Means per area unit e.g. percentage of land use class in total country area, number of trees per ha, volumes per ha, etc... These are ratio-estimators.
- Total estimates e.g. total forest area, volume...
- Sampling errors: standard errors of the ratio estimator as well as the variance of total estimates were calculated.

The final outputs of the calculations of study are:

- Area estimates of the different land use /forest type classes at different for the entire country, by ecological zone, by management, by ownership...
- Volume estimates;
- Biomass estimates (derived from volume equations)
- Carbon
- Diameter distribution

- Corresponding standard errors or sampling errors for given probability levels according to the particular sampling design.

5.5 Presentation of the results

5.5.1 What results to be presented?

In the final inventory report, mean estimates and total estimates are presented for all key parameters (area, volume, biomass, biodiversity).

Worth mentioning, is that the results should contain a certain level or degree of reliability as indicated by the standard error. The accuracy at which results are presented depends on the total error of these results. Accuracy of results should be of the same magnitude as the total error. Presentation of inventory results with accuracy much higher than their expected total errors is illusive and misleading. Emphasis must also be placed on consistency in accuracy of results.

Other statistical estimators such as variance or standard errors of the ratio should also be presented.

5.5.2 How to present results?

To vary presentation results may be presented as tables, pie charts or histograms.

6 Bibliography

Avery T.E., Burkhart H.E. 1994. *Forest Measurement*. Fourth Edition. McGraw-Hill, New York. 408 p.

Branthomme A. 2003. *Inventaire forestier du Cameroun : Manuel de terrain*. FAO/FRA, Rome.

Box G. E.; Hunter w.g. & Hunter J.S. 1978. *Statistics for Experimenters. An Introduction to Design, Data Analysis, and Model Building*. John Wiley & Sons. N.Y. 653p.

Cochran w.G. 1977. *Sampling Techniques*. Third Edition. John Wiley & Sons. N.Y. 428p.

Freese F. 1962. Elementary Forest Sampling. USDA Agriculture Handbook no. 232. 91 p.

Loetsch F. & Haller K.E. 1973. Forest Inventory volume 1. Third Edition. BLV VERLAGSGESELLSCHAFT MUNCHEN BERN WIEN. 436 p.

Loetsch F. Zohrer f. & Haller K.E. 1973. Forest Inventory. Volume 2. Third Edition. BLV VERLAGSGESELLSCHAFT MUNCHEN BERN WIEN. 469 p.

Rondeux J. 1993b. *La Mesure des Arbres et des Peuplements Forestiers.* Les Presses Agronomiques de Gembloux. 521 p.

Saket M., Altrell D., Branthomme A., Vuorinen P. 2002**.** *FAO's approach to support national forest assessments for Country Capacity Building. Background paper 6.6 to the Expert Consultation on Global Forest Resources Assessments – Linking National and International efforts.* Kotka, Finland, 1-5 July 2002.

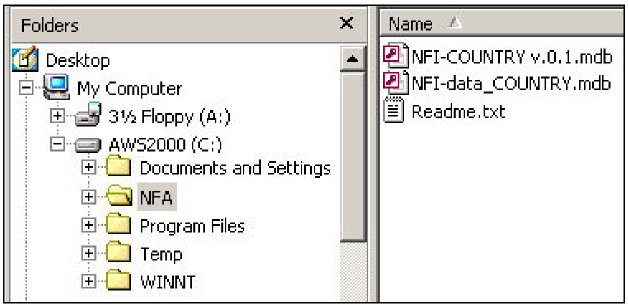
Scheaffer R.L.; Mendenhall W. & Ott L. 1990. *Elementary Survey Sampling*. Fourth Edition. Duxbury Press. 390 p.

Annex 1: NFA DATABASE APPLICATION - Getting started

Database file location

The two database files "*NFI-data_<COUNTRY>.mdb*" and "*NFI-<COUNTRY> v.x.x.mdb*" should be stored in a folder named "NFA" directly under C:\ (i.e. in C:\NFA\ see Figure 9).

Figure 9: Location of database files



The fixed location of the database files is necessary as the database application is linked to the dynamic tables in the "data" database and therefore need to know where the "data" database is located. In the database application, the location of the "data" database is set to the C:\NFA\ folder.

Opening the NFA database application

The user of the NFA database is managing the data through the application database. To open the database application the user double-clicks on the *NFI-<COUNTRY> v.x.x.mdb* file.

When starting the database application the user must enter a password in a dialogue box to access the application (see Figure 10).

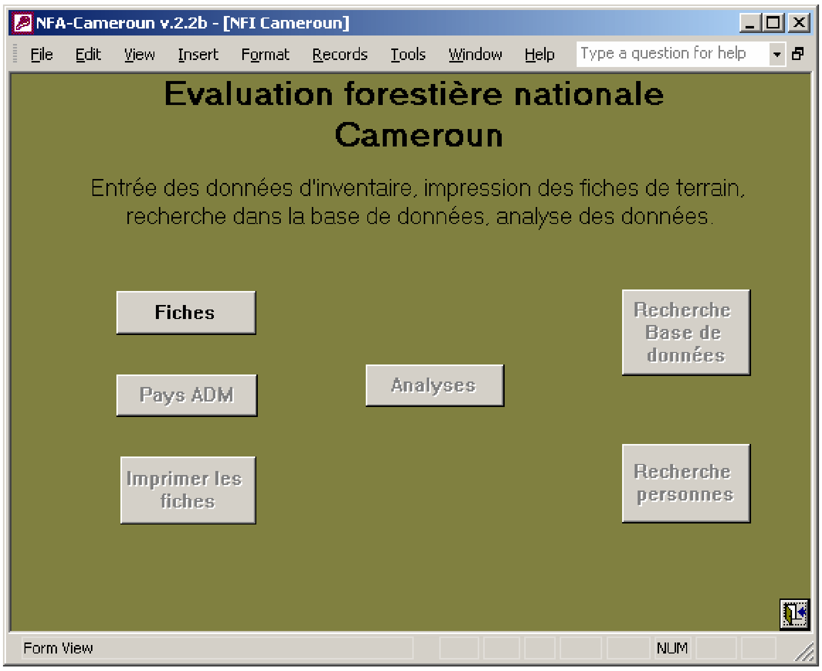
Figure 10: Log-on dialogue box

The data are managed through forms and the user navigates in the database by selecting different forms according to required operation.

Selecting database operation

After logging on with the password an initial form appears (11) where the user selects what kind of operations he/she wants to carry out.

Figure 11: Database operation selection form



The efforts in developing the NFA database application have initially been focused on wise data storage and on facilitating the input of primary data, why until now the only activated section in the database application is the section on "*Field Data*".

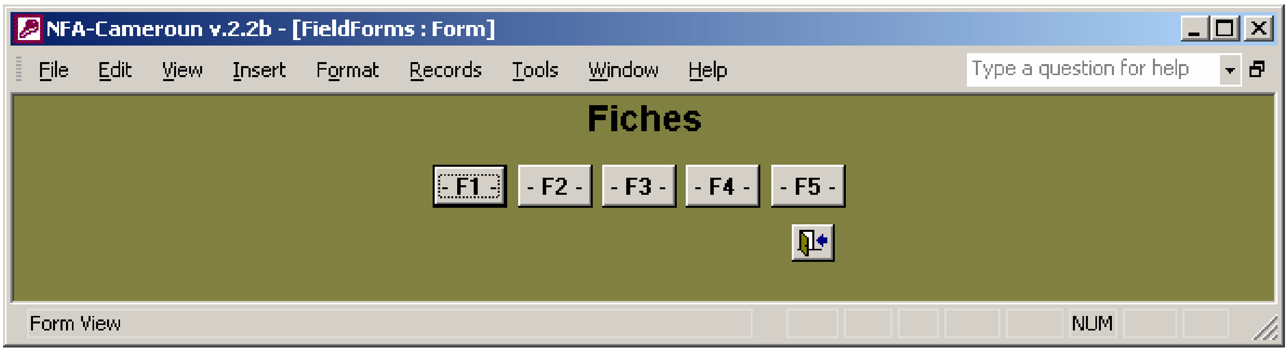
<u>Field data entry and management</u>

- **Selecting field form**

Through the "*Field Data*"/""Fiches" button in the *Database operation selection form* (**Figure 11**) the user accesses a form with a button for every corresponding field form (Figure 12).

In Figure 12 the *Field Form selection form* shows six [-F -] buttons, which represent the five main field forms employed for the field data collection. The user selects the button that corresponds to the field form he/she wants to edit or view.

Figure 12: Field forms selection interface



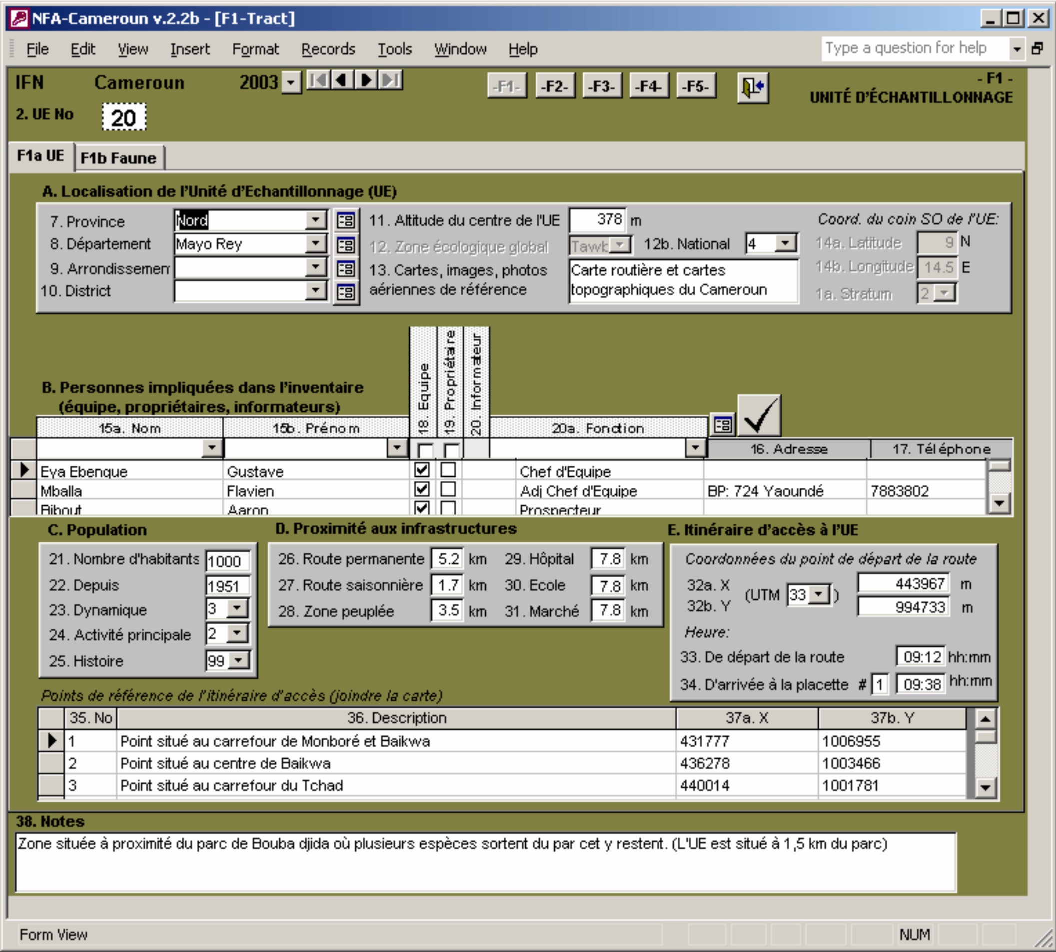
- **Field data entry and editing**

The design of the forms for adding/editing field data in the database application follows the same structure than the field forms employed for the data collection in the field. This facilitates the entry of field data into the database application since the way to register field data as given in the field manual is similar.

The following figures (Figure 4 to Figure 18) illustrate the forms used to add/edit the NFA field data. Further details on how the fill in the forms are given in Field Manual.

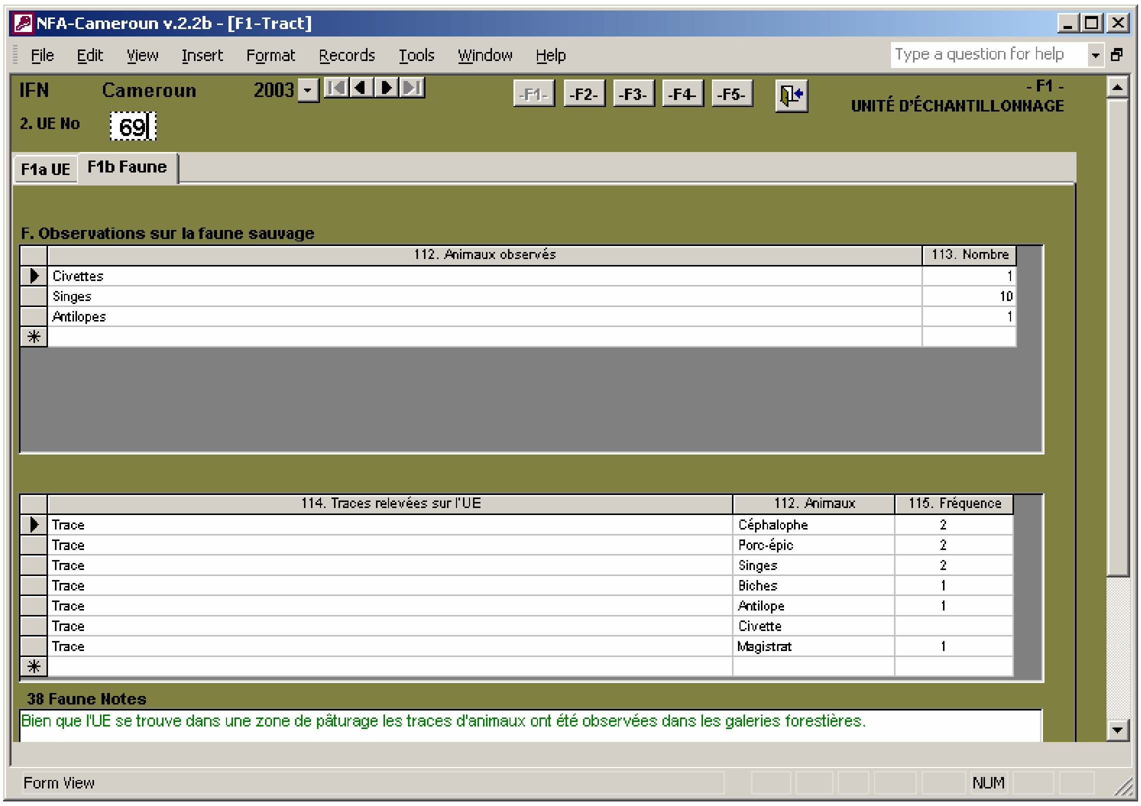
<u>Field form F1-Tract</u>

Figure 13: F1 - Tract attributes Form



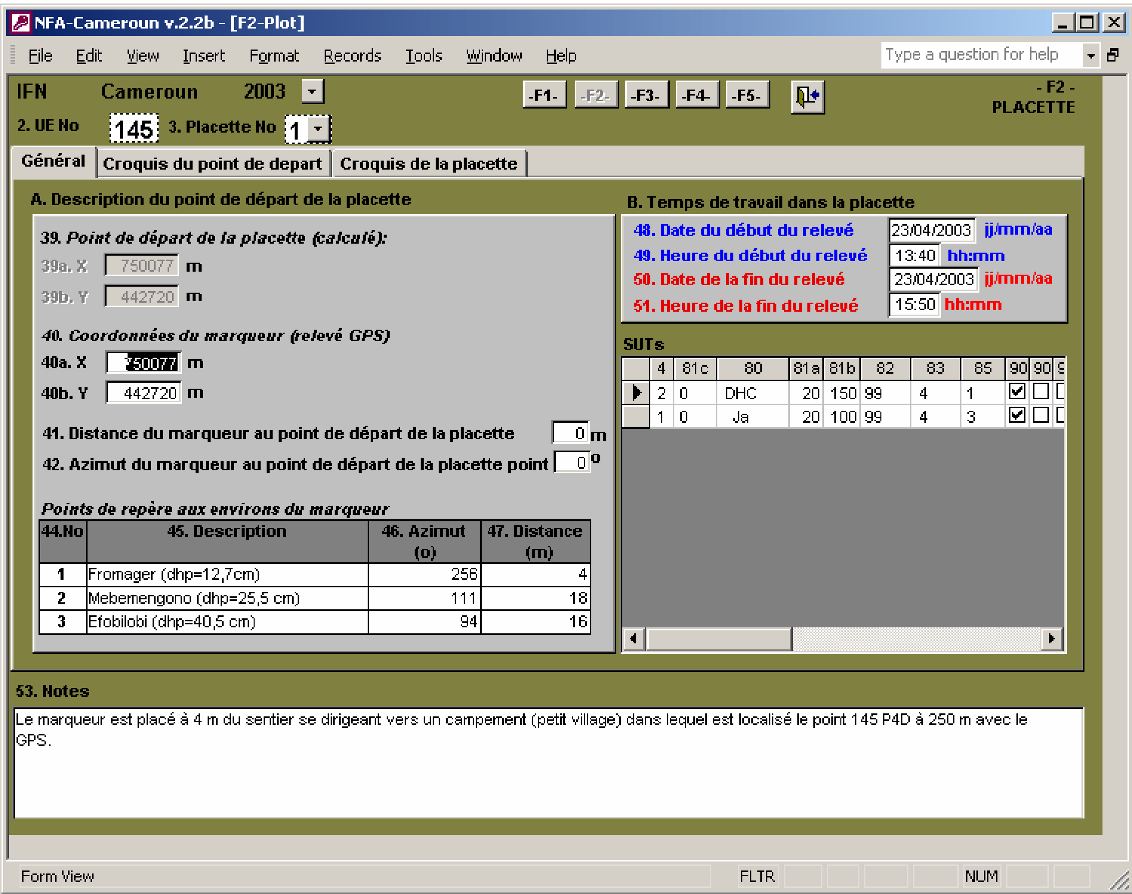
Field form F1b-Tract- Wildlife

Figure 14: F1 - Tract Observation on Wildlife attributes Form



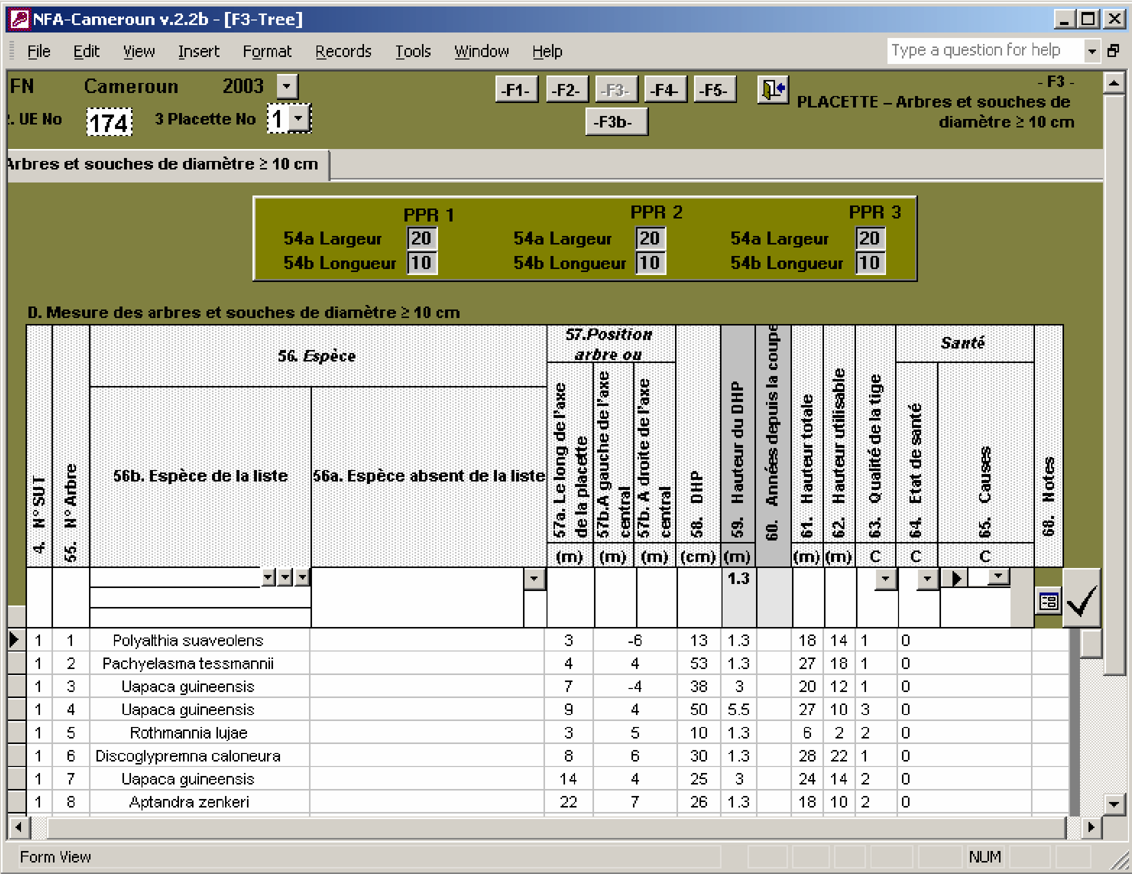
Field form F2-Plot

Figure 15: F2 - Plot attributes form



Field form F3a-Trees attribute data

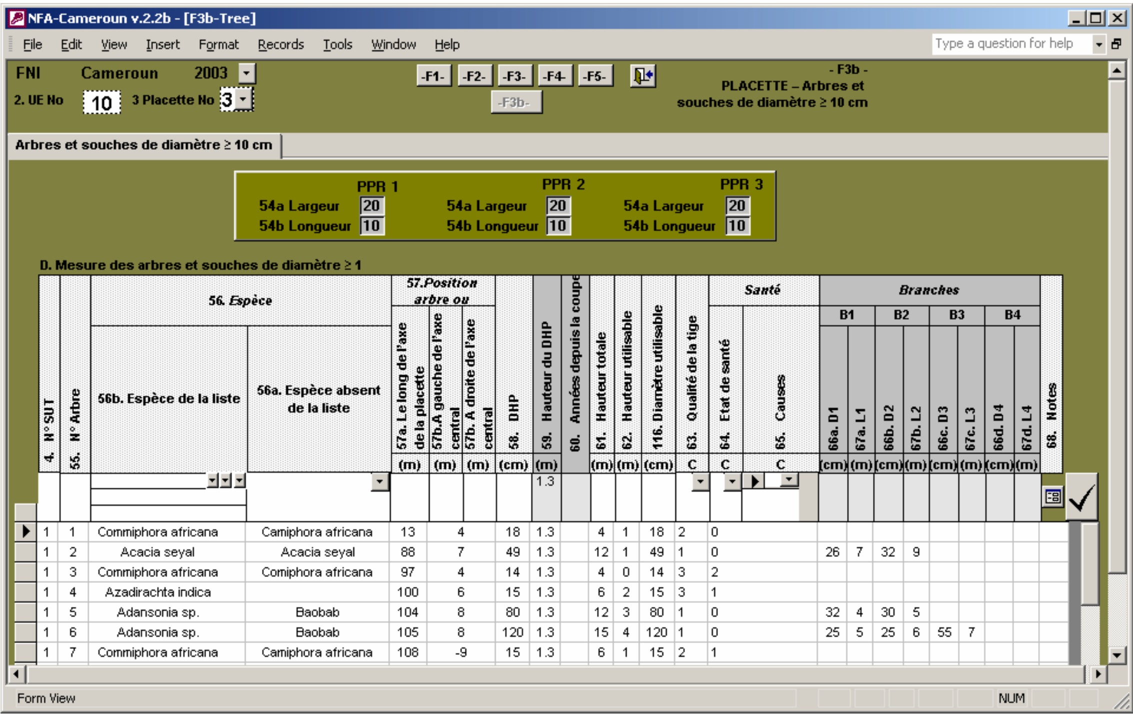
Figure 16: F3a – Tree attributes form



4. N° SUT	55. N° Arbre	56b. Espèce de la liste	56a. Espèce absent de la liste	57a. Le long de l'axe de la placette (m)	57b.A gauche de l'axe central (m)	57b. A droite de l'axe central (m)	58. DHP (cm)	59. Hauteur du DHP (m)	60. Années depuis la coupe	61. Hauteur totale (m)	62. Hauteur utilisable (m)	63. Qualité de la tige C	64. Etat de santé C	65. Causes C	68. Notes
								1.3							
1	1	Polyalthia suaveolens		3		-6	13	1.3		18	14	1	0		
1	2	Pachyelasma tessmannii		4		4	53	1.3		27	18	1	0		
1	3	Uapaca guineensis		7		-4	38	3		20	12	1	0		
1	4	Uapaca guineensis		9		4	50	5.5		27	10	3	0		
1	5	Rothmannia lujae		3		5	10	1.3		6	2	2	0		
1	6	Discoglypremna caloneura		8		6	30	1.3		28	22	1	0		
1	7	Uapaca guineensis		14		4	25	3		24	14	2	0		
1	8	Aptandra zenkeri		22		7	26	1.3		18	10	2	0		

Field form F3b-Tree attribute data including branches

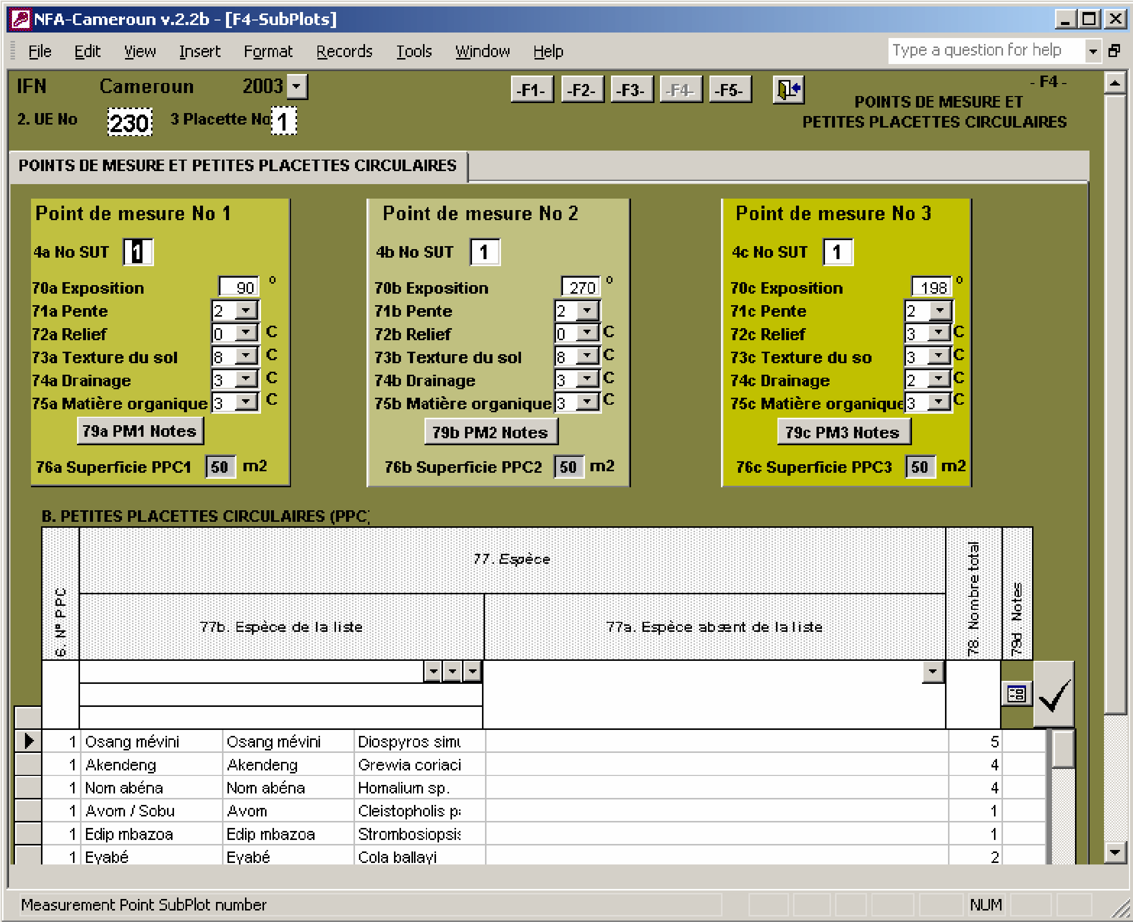
Figure 17: F3b – Tree attributes form including branches form



4. N° SUT	55. N° Arbre	56b. Espèce de la liste	56a. Espèce absent de la liste	57a. Le long de l'axe de la placette (m)	57b.A gauche de l'axe central (m)	57b. A droite de l'axe central (m)	58. DHP (cm)	59. Hauteur du DHP (m)	60. Années depuis la coupe	61. Hauteur totale (m)	62. Hauteur utilisable (m)	116. Diamètre utilisable (cm)	63. Qualité de la tige C	64. Etat de santé C	65. Causes C	66a. D1 (cm)	67a. L1 (m)	66b. D2 (cm)	67b. L2 (m)	66c. D3 (cm)	67c. L3 (m)	66d. D4 (cm)	67d. L4 (m)	68. Notes
								1.3																
1	1	Commiphora africana	Camiphora africana	13		4	18	1.3		4	1	18	2	0										
1	2	Acacia seyal	Acacia seyal	88		7	49	1.3		12	1	49	1	0		26	7	32	9					
1	3	Commiphora africana	Comiphora africana	97		4	14	1.3		4	0	14	3	2										
1	4	Azadirachta indica		100		6	15	1.3		6	2	15	3	1										
1	5	Adansonia sp.	Baobab	104		8	80	1.3		12	3	80	1	0		32	4	30	5					
1	6	Adansonia sp.	Baobab	105		8	120	1.3		15	4	120	1	0		25	5	25	6	55	7			
1	7	Commiphora africana	Camiphora africana	108		-9	15	1.3		6	1	15	2	1										

F4-Measurement Points & Subplots

Figure 18: F4a – Measurement point and sub-plot attributes form



F5-Land Use Section

Figure 19: F5 – Land Use Section (LUS) general attributes form

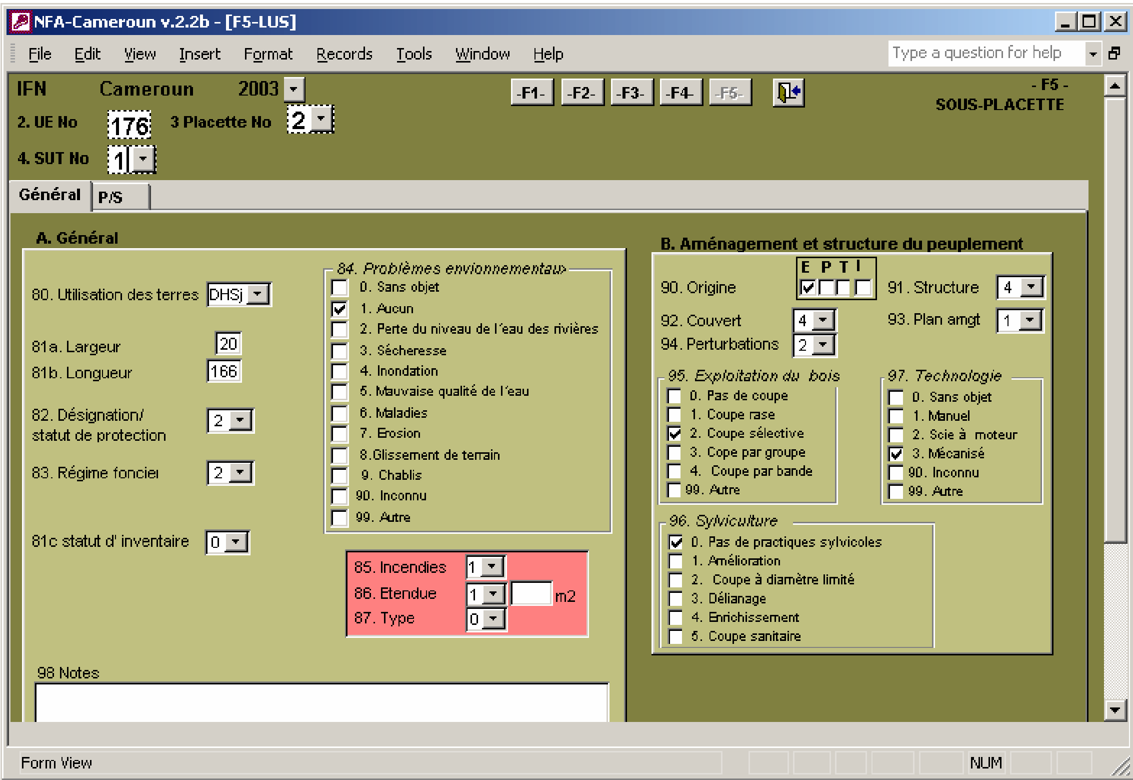


Figure 20: F5 – Land Use Section -Product and Service attributes form