CHAPTER 3- METHODOLOGY

3.1. Introduction

The data collection was conducted under the leadership of the Ministry of Agriculture, according to a sampling methodology with representativeness at the level of agro-climatic regions and "prefectures", dating back to 1983, when the last national agricultural survey was held. Since then, smaller agriculture surveys were conducted in successive years, with representativeness at the prefecture level.

These small surveys stopped in 1992 because of insecurity in the country. During the period of emergency that followed the genocide of 1994, it was imperative to roughly estimate the food availability and nutritional needs of the population in order to program food aid to supply in the short term. That is when the crop assessment campaigns started and this continues every season until today.

The agricultural production figures used so far are always provided by the crop assessment campaigns conducted shortly before the end of each growing season.
Several other surveys have been conducted on the rural sector but none has met the information needs either in terms of quality or in terms of representativeness. We can already cite surveys conducted between 1999 and 2002, by the Food Security Research Project (FRSP) using a representative sample of former prefectures, small farm surveys conducted in 2005 using a sample of 1,440 households and the 2006 which used a sample of 1,704 households

A broad-based survey was therefore necessary to ensure greater representativeness in space and time, observing a larger number of crops and providing a database to meet the agricultural statistics demand in the new context of decentralization, which enshrines districts as a starting point for development and planning. The introduction of new cash crops through horticultural activities deserves also to benefit from the monitoring provided by the survey. Hence this was the rationale of the National Agricultural Survey 2008 (NAS 2008).

3.2. Survey scope

The survey covers all rural areas of the country and some areas of Kigali city concerned by the agro-pastoral activities. The survey was conducted during the two agricultural seasons 2008A and 2008B which cover virtually the entire agricultural year. The same sample of households was visited during each growing season.
Since main users of agricultural statistics wished to have data at the level of districts, the following have been considered as key domain for NAS 2008.

3.3. Sampling of NAS 2008

3.3.1. Recent developments in the administrative structure of Rwanda

In 2002, Rwanda was divided into 11 provinces and Kigali City, 106 districts and 1545 Sectors. For purposes of the General Census of Population and Housing (GPHC/ RGPH) in 2002, each sector was divided into *Enumeration Areas* (EAs) with an average size of 227 households. Thus, Rwanda was

divided in 7727 EAs. These enumeration areas were well defined and identifiable on field, so that no EA overlapped on another.

In 2006, the number of decentralized administrative structures was revised downwards. The number of provinces was reduced to 5 (including Kigali City), the number of districts and sectors were reduced to 30 and 416 respectively.

3.3.2. Basic sampling units

The data for developing the sampling plan of the National Agricultural Survey 2008 come from the RGPH 2002 mapping and census results. The agricultural module that was administered to all households helped to identify agricultural households.

Information extracted from this module was used to develop a sample frame for the National Agricultural Survey. To select the primary sampling units (PSU), the number of agricultural households identified in RPGH 2002 was used as a measure of the size of the enumeration area (EA).

The unit of observation was the agricultural household. This was defined as the household where at least one member was engaged in any of the following; agricultural activities , livestock , fisheries , forestry or bee-keeping. A form for listing was used to identify this type of household. The unit of analysis was the holding of agricultural household. The agriculture sample frame consists of all agricultural households residing in the enumeration area.

3.3.3. Primary Sampling Unit (PSU)

At the first stage, the primary units of sampling (PSUs) constituted the sample frame of survey. The PSUs were geographical areas with clearly identifiable boundaries so that the listing of households was conducted by an enumerator during a fixed period of time. The sample EAs were drawn using probability proportional to size (PPS), the size of each PSU is the number of households. To ensure efficiency of the sample, it was expected to draw 840 EAs across the country.

3.3.4. Secondary Sampling Units (SSU)

The units that constitutes the frame in the second stage - in a two stage sampling - called Secondary Sampling Units (SSU) were formed by all agricultural households residing in the sample EA , from which a number of households were selected. The number of sample households by EA was determined such that the workload enabled the enumerator to complete the collection work on time. 12 agricultural households by EA were selected for data collection.

The selection was made so that all types of agricultural farmers, all types of crops and livestock and all cultures were represented in the sample.

3.3.5. The selection of sample households

After the selection of 840 primary units corresponding to the enumeration areas established during RGPH (2002) and using the principle of probabilities proportional to size, listing within sample EAs was done in order to establish the updated list of households residing in the EA. This operation led to identification of agricultural households. These lists of agricultural households helped for random selection of 15 agricultural households by EA with equal probability. Among these 15 households, 12 participated in interviews and 3 agricultural households served as replacements in case of failure or refusal to be surveyed by the other households.

3.3.6. The stratification of the NAS 2008

The first stage of stratification was at the district level in order to have the survey estimates at the district level. The districts of Kigali City (Nyarugenge, Gasabo, Kicukiro) were grouped into one stratum because of the small number of agricultural EAs in each district. Thus, at the first stage of stratification, there are 28 independent strata which corresponded to the domains of analysis: the City of Kigali and each of the 27 remaining districts.

Furthermore, Rwanda was divided into ten bio- climatic zones. Considering that the agricultural and livestock production was closely linked to bio climatic zones, implicit stratification following the "bio-climatic zones" increased the efficiency of sampling. Implicit stratification consisted of ordering the EAs in each stratum by the criterion of choosing before the selection of the PSU.

The bio-climatic zones are defined as follows:

(01)   Cyangugu countryside
(02)   Banks of Lake Kivu
(03)   Cones and high volcanic planes
(04)   Congo Nile Ridge
(05)   Ridges and plateau bordering the Savanah of the east
(06)   Buberuka Highlands
(07)   Mayaga and Bugesera
(08)   Plains of Bugesera
(09)   Central Plateau
(10)   Savannah of east and central Bugesera

Each agricultural EA was classified into one of the ten bio climatic zones. The analysis of agricultural sample frame showed that most districts (19) had two bio climatic zones, 6 districts had three bio climatic zones, 2 districts (Musanze and Nyamasheke) had 4 zones and 3 districts were in a single agro climatic zone.

3.3.7. Sample size and precision of estimates

The desired size for the sample was determined by taking into account several factors, including the three main ones; the degree of reliability expected of survey estimates, cost and and the efficiency of the sample. However, the precision of estimates from the survey depended on the effective control of sampling and non-sampling errors. In consideration of financial and operational constraints, and in order to have reliable estimates at district level, it was suggested to select 840 EAs at the first stage. This proposed sample was divided into 4 sub-samples of 210 EAs for each in preparation of subsequent light surveys (including the Survey of post harvest crop production, etc..) and according to resources that could be mobilized.

In the second stage, 12 households were selected in each sample EA. For the National Agricultural Survey 2008, there were a total of 10,080 sample households.

Sample distribution

Table 3.3.1. Distribution of the agricultural base and distribution of sample EAs per district

District	Number of EAs	Total agricultural households	% Agricultural households	Prorate	Adjustment of the number of EAs
VK (Gasabo, Kicukiro, Nyarugenge)	214	45800	3.0%	25	**28**
BUGESERA	230	56329	3.6%	31	**32**
BURERA	300	66683	4.3%	36	**32**
GAKENKE	328	71290	4.6%	39	**36**
GATSIBO	259	61027	3.9%	33	**32**
GICUMBI	331	69739	4.5%	38	**36**
GISAGARA	250	60630	3.9%	33	**32**
HUYE	227	51608	3.3%	28	**28**
KAMONYI	244	53431	3.4%	29	**28**
KARONGI	260	54103	3.5%	29	**28**
KAYONZA	202	45864	3.0%	25	**28**
KIREHE	207	50375	3.2%	27	**28**
MUHANGA	263	54454	3.5%	29	**28**
MUSANZE	265	58375	3.8%	32	**32**
NGOMA	221	51209	3.3%	28	**28**
NGORORERO	277	63653	4.1%	34	**32**
NYABIHU	246	55738	3.6%	30	**32**
NYAGATARE	238	51464	3.3%	28	**28**
NYAMAGABE	279	58342	3.8%	32	**32**
NYAMASHEKE	343	63715	4.1%	35	**32**
NYANZA	203	47849	3.1%	26	**28**
NYARUGURU	236	49586	3.2%	27	**28**
RUBAVU	212	47008	3.0%	25	**28**
RUHANGO	233	52425	3.4%	28	**28**
RULINDO	237	53452	3.4%	29	**28**
RUSIZI	299	59054	3.8%	32	**32**
RUTSIRO	268	55112	3.6%	30	**28**
RWAMAGANA	196	42583	2.7%	23	**28**
Total	7068	1550898	100.0%	840	**840**

Proportional distribution was adjusted in order to obtain a minimum of 28 EAs and a maximum of 36 EAs. The number of EAs was rounded to a multiple of 4 for identifying 4 sub samples for subsequent surveys according to budget availability.

Table 3.3.2. Distribution of sample ZDs and agricultural households for NAS and for a sub sample of 25%

District	NAS Sample		25% of NAS sample	
	Number of Sample EAs	Number of sample households	Number of Sample EAs	Number of sample households
VILLE DE KIGALI (Gasabo, Kicukiro, Nyarugenge)	28	336	7	84
BUGESERA	32	384	8	96
BURERA	32	384	8	96
GAKENKE	36	432	9	108
GATSIBO	32	384	8	96
GICUMBI	36	432	9	108
GISAGARA	32	384	8	96
HUYE	28	336	7	84
KAMONYI	28	336	7	84
KARONGI	28	336	7	84
KAYONZA	28	336	7	84
KIREHE	28	336	7	84
MUHANGA	28	336	7	84
MUSANZE	32	384	8	96
NGOMA	28	336	7	84
NGORORERO	32	384	8	96
NYABIHU	32	384	8	96
NYAGATARE	28	336	7	84
NYAMAGABE	32	384	8	96
NYAMASHEKE	32	384	8	96
NYANZA	28	336	7	84
NYARUGURU	28	336	7	84
RUBAVU	28	336	7	84
RUHANGO	28	336	7	84
RULINDO	28	336	7	84
RUSIZI	32	384	8	96
RUTSIRO	28	336	7	84
RWAMAGANA	28	336	7	84
TOTAL	840	10080	210	2520

3.4. Weighting procedures

To make sample estimates from the Agricultural Surveys representative for all agricultural households of the country, it was necessary to multiply the data by a sampling weight, or expansion factor. The basic weight for each sample agricultural household was equal to the inverse of its probability of selection. This probability was obtained by multiplying the probabilities at each stage of selection.

$$p_{hi} = \frac{z_h \times M_{hi}}{M_h} \times \frac{m_{hi}}{M'_{hi}},$$

where:

p_{hj} = probability of selection for the sample agricultural households in the ith sample EA in stratum h
z_h = number of sample EAs in stratum h

M_h = total number of agricultural households in the frame (based on RGPH 2002 / Rwanda Census) for stratum h
M_{hi} = total number of agricultural households in the frame for the ith sample EA in stratum h
$m_{hi} = 12$ = number of sample agricultural households selected in the ith sample EA in stratum h
M'_{hi} = total number of agricultural households listed in the ith sample EA in stratum h

The basic sampling weight, or expansion factor, was calculated as the inverse of this probability of selection.

$$W_{hi} = \frac{M_h \times M'_{hi}}{z_h \times M_{hi} \times m_{hi}},$$

where:

W_{hi} = basic weight for the sample agricultural households in the ith sample EA in stratum h
m_{hi} = 12 = number of sample agricultural households originally planned to be selected in the ith sample EA in stratum h

To take into account the noninterview rate during data collection, expansion factor is adjusted as follows:

$$W'_{hi} = W_{hi} \times \frac{m_{hi}}{m'_{hi}},$$

where:

W'_{hi} = final weight of sample household in the ith sample EA in stratum h;

m_{hi} = 12 = number of sample agricultural households originally planned to be selected in the ith sample EA in stratum h

m'_{hi} = total number of actually interviewed sample households in the ith sample EA in stratum h

3.5. Data collection

The data collection was done using questionnaire sheets which were filled by enumerators according to a harmonized calendar in all selected EAs. Each enumerator had to work in two EAs and visit 12 households per EA, i.e 24 households per enumerator. Through the questionnaire sheets, it was possible to collect a large quantity of figures that made it possible to measure several indicators related to farming and livestock breeding within sample households. To fill questionnaire sheets, it was necessary to visit each sample households several times. The questionnaire sheets used are presented in Annex 2.

In addition to the collection sheets in the NAS 2008, the concern of measurement of accuracy was satisfied through the distribution to enumerators and heads of households standard measurement equipments. This was done in order to break the tradition of approximation of quantities and distances as used in previous surveys. Thus, each enumerator was equipped with a measuring tape to get data regarding the dimensions of fields and farms, a spring balance to measure the weights of harvested crops and a calculator. Each household received a bucket calibrated in liters to measure liquid products, a spring balance and a sack to measure the weight of solid products and a pencil to record the amounts collected regularly on an appropriate register.

Regarding supervision of data collection, a statistician appointed in each district by NISR assumed the role of coordinator of field activities. For this, all the statisticians received transport facilitation from NISR with a vehicle for monitoring during at least 7 days per month per district. District statisticians and survey controllers at district level also received monthly telephone cards worth 20,000 RWF in order to maintain contact with field workers.

The enumerators had previously received in the beginning of agricultural season, a calendar for filling questionnaires sheets. At specific times, under the guidance of the survey management team, the enumerators proceeded to rally district by district, to deliver and check the filled questionnaires.

At national level, a supervision team carried out missions to intervene on the ground and to solve any problems of discipline and disorganization.

The collection staff on site was composed of:

- 428 enumerators working in the districts (14 to 18 per district);
- 56 controllers ie 2 controllers by District / stratum.

3.6. Data processing

3.6.1. Data Entry

Upon the closing of fieldwork of season 2008A, a large volume of data was available for entry. This required a consistent supply and a significant number of staff input to recruit and employ. Unfortunately, organizational and financial difficulties arose and led to a late start of data entry.

For data entry operations, a computer program was developed using a CSPro statistical software application with a questionnaire on each sheet. A training session was organized on this data entry program and on the nature and extent of work to do. In total, this operation mobilized 184 data entry clerks, 10 controllers, 3 checkers, 3 supervisors of coding and 92 computers. A first team of data entry clerks, controllers and checkers worked in the morning and was be relayed by a second team to work in the evening.

Data entry was done in two steps:

a) Data entry of the 2008A season: from June to July 2008

The preparation of this step consisted of checking completed forms and returning to site in case of improperly completed sheets, for correction. From 23/06/08 to 12/07/08, three field supervisors skilled in surveys conducted a mission at the headquarters of the NISR in order to help in organizing, supervising and coaching data entry clerks. They played the role of controllers and checkers of coding during the data entry period.

b) Data entry of the 2008B season: from September to November 2008

As it was not possible to return to field records improperly completed during the 2008B season at the end of the field work, special instruction was given to field workers to undertake collectively a verification and correction of errors in questionnaires before delivering them to the headquarters of NISR. This instruction was observed by all fieldworkers. Subsequently, the data entry was done with the support of the same controllers and checkers of coding.

3.6.2. Data cleaning and processing

A computer statistician consultant[1] was recruited for the cleaning work and data processing for NAS 2008. The methodology followed in cleaning and processing the data is contained in the document *«Plan méthodologique pour l'apurement et le traitement des données»* formulated by the consultant. The consultant had to monitor compliance with the codification of collection sheets and make the necessary corrections.

---

[1] Mr. Jean-Marie Vianney Sehene

3.7. Data analysis

An international consultant has been recruited to carry out the analysis of the NAS 2008[2]. The analysis covered the demographic and social characteristics of agricultural farmers, farms characteristics, agricultural practices and crop production, livestock practices and production, fishery, aquaculture and beekeeping practices, forestry practices and income, as well as food stocks and nutrition of agricultural households. At this end, the following steps were followed:

3.7.1. *Consultation of documents and technical and methodological survey reports*

The first task was the reading of all technical and methodological survey reports and with the members of NISR technical team who were in charge of the design and the implementation of the survey. That step was very important as it allowed better understanding of how and why the survey was designed, planned and implemented; it was also an opportunity to understand possible problems that may have occurred. This therefore helped for the right evaluation of the resulting data quality.

The reading also covered reports and documents containing information on agricultural development policy prevailing in Rwanda. Indeed, chosen survey modules were targeting to provide benchmark data in order to help decision making vis-à-vis the required strategy in the framework of the national agricultural policy. Hence, it is through the analysis of data obtained in regards of the objectives of that policy that the data analysis has been conducted.

3.7.2. *Consultations with partners*

In order to achieve a good understanding of the way the survey was designed and conducted, consultations with concerned partners were undertaken. In addition to concerned NISR technical staff with whom permanent discussions were maintained, the consultations also included other key data producers and users in Rwanda as well as survey field staff. In this way, everybody contributed to the improvement of the data analysis exercise. The final outcome is therefore considered as a team work output. At the same time, it responded to the expectation of users.

To this end, the preliminary results of the survey were shared and discussed together with key required partners (through a technical meeting organized for this purpose). Observations raised from that meeting were considered for the revision and finalization of data tabulation. It is at that point that the actual data analysis was initiated.

---

[2] Mr. Vincent Ngendakumana

3.7.3. Verification of the preliminary results of the 2009 NAS of Rwanda

It was important to proceed with an in-depth verification of the preliminary survey results for the following main reasons. Firstly, in order to ensure that any needed information was produced in the required format; secondly, to ensure that the data produced was reliable, consistent and coherent; and lastly, in order to make possible data corrections and/or adjustments (with the help of the computer specialist recruited for data cleaning and processing). Particular attention was paid to the right use and respect of standard and international related concepts, definitions and classification of agricultural products.

3.7.4. Collection and compilation of agricultural routine data

Frequently, 2008 NAS results were compared to other preexisting routine data (according to their availability and reliability). For that reason, it was important to proceed first with routine data compilation from surveys conducted in the past by other partners (especially agricultural production, yield and area, and livestock numbers and production MINAGRI). A relevant format for the compilation of data was used. The obtained data was used as baseline in order to evaluate the reliability of NAS results. In this way, some data, like those on cash crops (coffee and tea) traditionally produced by their respective systems (OCIR-Café and OCIR-Thé), revealed themselves to be well in line with those from of the NAS 2008. Discrepancies observed at the level of other data from other sources was explained mainly by gap and weaknesses noticed with methodological approaches used. Therefore, this cannot make questionable, the reliability of the NAS 2008 results published herein.

On the other hand, usage of other parameters (like the technical conversion factors, nutritional factors) for the generation of supplementary data allowed better appreciation of the relevancy of NAS 2008 results. It is by doing so that the conversion of the agricultural productions in nutritional equivalent allowed the conclusion that they were well of a very reasonable level and that they deserved to be accorded acceptable accuracy.

3.7.5. Analysis of the NAS 2008 results and preparation of the report of analyzed results

The themes of analysis of the results were previously agreed upon and were based on the questionnaire's modules. Prior to the analysis work, a detailed plan of the report was established and agreed upon and the contents of different chapters and sections of the report were thereafter elaborated.

The analysis was done after the data was extensively verified and approved. The analysis work consisted of explaining (in simple and relevant terms) and pointing out prominent facts emerging from the survey results. A geographical comparative analysis was made at district and provincial levels of the country. Whenever needed and possible, the data analysis was supported by graphic and/or cartographic representation of important phenomena. For some particular cases, and as mentioned in the section 3.7.4, one had to resort to supplementary information in order to generate specific additional indicators.