FAO STATISTICAL DEVELOPMENT SERIES

10

Multiple frame agricultural surveys

Volume 2

Agricultural survey programmes based on area frame or dual frame (area and list) sample designs

Food and Agriculture Organization of the United Nations

FAO STATISTICAL DEVELOPMENT SERIES

10

Multiple frame agricultural surveys

Volume 2

Agricultural survey programmes based on area frame or dual frame (area and list) sample designs

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS
Rome, 1998

M-77
ISBN 92-5-104074-5

FOREWORD

One of the major tasks of FAO is to improve national systems for the collection of agricultural statistics. For this purpose, it is particularly important to promote the application of the most appropriate statistical methods and practices for current national agricultural surveys and censuses.

The Statistics Division of FAO published, over the years, two main books in the field: *Sampling methods and censuses*, Zarkovich (1965), and *Sampling methods for agricultural surveys,* Kish (1989). The FAO Statistics Division's contributions to the methodological work on agricultural surveys and censuses are also reflected in P.V. and B.V Sukhatme's *Sampling theory of surveys with applications* (1970).

This is the second volume of a two-volume book on *Multiple frame agricultural surveys*. It was prepared as a contribution to the study of statistical methods and practical procedures necessary for planning and conducting agricultural survey programmes. It is an introduction to area and multiple frame probability sampling methods and their application, with a special view to meeting the requirements and constraints faced by developing countries.

Area and multiple frame designs (that usually consist of a dual - *area* and *holding* - frame) provide, in a large range of countries, the statistical foundation of a variety of national agricultural surveys: general purpose periodic agricultural production and livestock surveys; objective crop yield surveys; cost of production surveys; chemical use (fertilizers and pesticides) surveys; land use surveys; labour force surveys; environmental surveys; and social and economic surveys of the agricultural sector.

The first volume of this book, published in 1996, provides a general classification of alternative agricultural survey designs with an indication of their respective advantages and limitations. It examines several aspects which have to be considered to establish and conduct a periodic agricultural survey programme based on dual frame sampling methods, i.e. the probability selection and estimation methods, the survey organization, the equipment and materials needed, the data collection, the summarization and processing procedures, as well as the preliminary analysis of the survey results. The volume also includes a detailed description of a category of survey designs considered especially useful for developing countries.

This second volume presents the survey methodologies of area and multiple frame agricultural survey programmes currently used in a large range of countries. It thus provides actual examples of the application of the survey methods presented in the first volume. The purpose is to allow the reader to make the necessary comparisons and to facilitate the application of the most appropriate statistical methods and practices.

Ladislav Kabat
Director
Statistics Division

ACKNOWLEDGEMENTS

The FAO two-volume publication *Multiple Frame Agricultural Surveys* was written by Alvaro González-Villalobos, Senior Statistician, jointly with Montie A. Wallace, FAO consultant.

The agricultural survey methodological descriptions presented in this volume were prepared by the authors on the basis of publications and their knowledge of the survey designs involved. Some survey methodological descriptions were based mainly on national case studies prepared at FAO's request by the following experts: William H. Wigton, Zamir Beliu, and Altin Gjika (Albania), Osvaldo Stepancich (Argentina), Gil Silva, Mirane M. Carrilho and Elvira N. Pitaluga (Brazil), Jacqueline E. Mayda and Nanjamma S. Chinnappa (Canada), Robert Arcaraz and Jean-Claude Porchier (France), Gilberto Gálvez Cálix and Roberto Guevara Salinas (Honduras), Aldo Giovacchini, Paolo Ragni and Fabrizio Balli (Italy), Seghir Bouzaffour (Morocco), Michael A. Steiner and Ray Garibay (Nicaragua), Sharif Ahmad Khan (Pakistan), and Luis Ambrosio Flores and Francisco Javier Gallego Pinilla (Spain).

FAO wishes to acknowledge, as consultants and external reviewers, Mirane M. Carrilho for her contribution to chapter 11 and the general review of the final version of this volume; and Elisabetta Carfagna for her work on the agricultural survey methodological descriptions of the European countries.

FAO also wishes to thank other distinguished statisticians: Leslie Kish, R. Paul Moore, Larry A. Sivers, George A. Hanuschak, Robert C. Hale, Paul W. Blackwood, John Kovar, Mahamoud S.M. Nazif, Mostafa K. Nosseir, Josephine S. Wallace, Clarence Dunkerley and Hassan El Sheikh Elbashir for their support.

The editing and the preparation of the Index of Technical Terms was done by Laura Viaggio; Gianna Marciani-Politi, Annie Vilcocq and Susan Behravesh provided secretarial support for the publication of this volume.

CONTENTS

CHAPTER 2. ***CANADA***

CHAPTER 3. *HONDURAS*

CHAPTER 4. *UNITED STATES*

CHAPTER 6. *ARGENTINA*

CHAPTER 7. *MOROCCO*

CHAPTER 13. ***ITALY***

INTRODUCTION

National Agricultural Survey Programmes established to obtain reliable and timely basic data on the agricultural sector are based on one of three sampling survey statistical methods: *list sample designs*, in which the sampling units are the farms; *area sample designs*, in which the sampling units are small land areas (segments); and *multiple frame designs*, in which commonly a dual sampling frame is utilized so that an area sample in combination with a list sample is used.

Agricultural survey programmes based on area or multiple frame probability sampling methods should constitute, in a larger range of countries, the foundation of the agricultural statistical systems, in order to obtain reliable and timely basic data on the agricultural sector.

In fact, area or multiple frame probability sampling methods provide, in many countries, the statistical foundation of a variety of national agricultural survey programmes: general purpose current agricultural production and livestock surveys; objective crop yield surveys; agricultural structure surveys; cost of production surveys; chemical use (fertilizers and pesticides) surveys; land use surveys; labour force surveys; environmental surveys; and social and economic surveys of the agricultural sector.

Multiple frame sampling designs commonly use a dual sampling frame: an area frame that consists of small land areas (segments) and a list frame of agricultural holdings/holders. An *area sample survey*[1] designates a probability sample survey in which the final stage sampling units are land areas called *segments* and the selection probabilities are proportional to their area measures. The size measures used to select the segments (sampling units) are defined as a function of their area measurements. The usual type of size measure is the total area. On the other hand, a *list sample survey* designates a probability sample survey that is not an area sample survey. The usual type of sampling units and reporting units for list sample surveys are the agricultural holdings.

The dual frame sample designs can benefit from the advantages of both types of samples: the list sample is extremely efficient for estimating data of large holdings and holdings that produce rare items, while the area sample gives a better probability model that ensures a complete coverage and provides unbiased estimates.

This is the second volume of a two-volume book entitled *Multiple Frame Agricultural Surveys*. It was prepared as a contribution to the study of statistical methods and practical procedures necessary for planning and conducting agricultural survey programmes. It is an introduction to area and multiple frame probability sampling methods and their application, particularly methods applicable in developing countries. It also includes multiple frame designs that are appropriate for many developed countries.

---

[1] A note on the use of the term "area sample" follows this Introduction.

The two-volume publication constitutes the first comprehensive introduction to general purpose agricultural survey designs based on area or multiple frame sampling methods, including a methodological description of actual regular survey programmes conducted in a significant number of countries.

The first volume, published in 1996, provides a general classification of alternative agricultural survey designs with an indication of their respective advantages and limitations. It examines several aspects which have to be considered to establish and conduct a periodic agricultural survey programme based on dual frame sampling methods, i.e. the probability selection and estimation methods, the survey organization, the equipment and materials needed, the data collection, the summarization and processing procedures, as well as the preliminary analysis of the survey results. The volume also includes a detailed description of a category of survey designs considered especially useful for developing countries, i.e. the designs involving data collection through personal interviews with farmers that consist of a complete enumeration list frame, combined with an area frame in which the segments have recognizable physical boundaries.

This second volume presents the survey methodologies of area and multiple frame agricultural survey programmes currently used in a significant range of countries worldwide. It thus provides actual examples of the application of the survey methods presented in the first volume. The purpose is to allow the reader to make the necessary comparisons and to facilitate the application of the most appropriate statistical methods and practices.

The first part of this volume includes methodological descriptions of *general purpose* agricultural survey programmes based on area or multiple frame sampling methods. The current survey programmes based on multiple frame sampling methods conducted in Brazil, Canada, Honduras and the United States and the survey programmes based on area sampling methods conducted in Albania, Argentina, Morocco, Nicaragua, Pakistan, Colombia, Costa Rica, the Dominican Republic, El Salvador and Guatemala are presented. Since the area and multiple frame designs described correspond to *general purpose* agricultural surveys, they necessarily involve the data collection to be conducted through interviews with farmers. Indeed, as is well known, most of the agricultural characteristics (variables) studied are not amenable to measurement without conducting interviews with farmers.

The second part of this volume provides a detailed presentation of the direct expansion area sample estimators and their variances most commonly applied in general purpose agricultural survey designs as described in Part I. It also includes multiple frame estimators and the interpretation of the different types of area sample designs. It provides additional considerations and formulae to those given in Chapter 4 of Volume I. Other area sample estimators are included in the appropriate summaries.

The third part of this volume describes *special purpose* agricultural survey programmes in European countries which do not involve data collection through personal interviews with farmers and which are based on area sampling methods using square segments. The land use survey in France, the crop production survey in Italy and the crop area survey conducted in some regions of Spain are described. The special purpose surveys described are based on different area sampling methods from those used for general purpose agricultural survey programmes, having more restricted objectives.

Some of the country survey methodological descriptions presented in this volume were prepared by the authors mainly on the basis of case studies requested by FAO (Albania, Argentina, Brazil, Canada, France, Honduras, Morocco, Nicaragua, Pakistan and Spain). These case studies were summarized by the authors in such a way that the integrity and individuality of each report were preserved. The conclusions and considerations on future developments included in some of these country summaries were selected from the recommendations given in the original case studies.

The presentation in this volume of the agricultural survey methodological descriptions also fills a need for documentation of the work done so far in most of the selected countries. Knowledge of past activities allows for the repetition building on what was successful while avoiding previous errors and for improving the overall survey procedures. In most countries, a comprehensive and useful methodological description of their agricultural survey programmes is needed.

This second volume also contains an Index of Technical Terms, and a more comprehensive Bibliography.

NOTE ON THE TERM "AREA SAMPLING"

There are two commonly used meanings of the statistical term "area sampling", which often leads to misunderstanding among survey statisticians. They are the following:

Definition A (restricted meaning). An *area sample survey* designates a probability sample survey in which the final stage sampling units are land areas called *segments* and the selection probabilities are proportional to their area measures. The size measures used to select the segments (sampling units) are defined as a function of their area measurements - the most common type of size measure is the total area. The usual reporting unit in agricultural area sample surveys is the agricultural holding. On the other hand, a *list sample survey* designates a probability sample survey that is not an area sample survey. The usual sampling units and reporting units for list sample surveys are the agricultural holdings.

Definition B (general meaning). An *area sample survey* designates a probability sample survey in which, at least for one sampling stage, the sampling units are land areas.

- The survey methods based on area sampling following Definition A are of primary importance for agricultural statistics and related fields. Indeed, the segments (land areas) are very appropriate last stage sampling units to study important agricultural variables as is the case of crop areas or variables related with land areas. However, these area sampling methods are not of primary importance for non-agricultural surveys (survey designs on population, housing, businesses, health and many other fields).

- Definition A is mainly used in many, but not all, publications concerning agricultural survey methods. These methods generally involve an additional complexity in the estimation procedures, due to the use of very different sampling and reporting units.

- Definition B is very natural and appropriate to describe the technical meaning of the term. It is the definition used in the classical books on statistical survey methods and practices: namely, Cochran; Deming; Hansen, Hurwitz and Madow; Kish; Murthy; Sukhatme; and Yates. These books are, and have been for many years, the main reference and most widely used books for teaching survey statisticians around the world. It is most likely that this is one of the reasons why this definition of area sampling *is the more widely used among survey statisticians and probably even among agricultural survey statisticians*. The above-mentioned classical books on general survey methods and practices refer mainly to non-agricultural surveys, and do not include the survey procedures based on area sampling following the more restricted Definition A.

The definition of area sampling in its more restricted sense (Definition A), although probably less commonly used, is adopted in many agricultural survey books, and in particular in the present book. Indeed, the more restricted meaning of the term is adopted because it is well understood by a large number of agricultural statisticians and because a better term to define such an important category of agricultural survey designs has not yet been found.

PART I

GENERAL PURPOSE AGRICULTURAL SURVEYS BASED ON AREA OR MULTIPLE FRAME SAMPLING METHODS

As defined in Volume I, a *general purpose Current Agricultural Survey Programme* is a periodic (annual or seasonal), national (or large-scale), agricultural probability sample survey designed to obtain basic data for the agricultural sector. Survey estimates are required, it is assumed, for most of the following *variables*: crop areas (prepared, planted and harvested), crop yields (forecasted and achieved), crop production, livestock inventories, grain stocks, farming systems, cost of production, farm expenditures and social and economic characteristics of the agricultural holdings.

This first part includes methodological descriptions of selected general purpose agricultural survey programmes based on area or multiple frame sampling methods. The current survey programmes based on multiple frame sampling methods conducted in Brazil, Canada, Honduras and the United States and the survey programmes based on area sampling methods conducted in Albania, Argentina, Morocco, Nicaragua, Pakistan, Colombia, Costa Rica, the Dominican Republic, El Salvador and Guatemala are presented.

Since the area and multiple frame designs described correspond to *general purpose* agricultural surveys, they necessarily involve the data collection to be conducted through interviews with farmers. Indeed, as is well known, most agricultural characteristics (variables) studied are not amenable to measurement without conducting interviews with farmers.

CHAPTER 1

BRAZIL

GENERAL PURPOSE AGRICULTURAL SURVEY PROGRAMME BASED ON MULTIPLE FRAME SAMPLING METHODS (1986-1997)

1. THE SURVEY AREA

The general purpose Agricultural Survey Programme based on multiple sampling frame methods annually covers an important Brazilian agricultural region (the States of Paraná, Santa Catarina, Distrito Federal and São Paulo), an area of 550 000 km^2, that includes around one million agricultural holdings. The Survey Programme provides annual estimates of planted and harvested crop areas, areas intended for harvest, potential and actual yields and production for the major summer crops, livestock estimates, characteristics of the holdings and farming systems. It was designed and implemented by the Brazilian Geography and Statistics Institute (IBGE) in cooperation with the Brazilian Space Research Institute (INPE). After the programme was designed and implemented in the State of Paraná, the World Bank provided financial assistance for a three-year period.

Geographic Subdivisions. In Brazil, the *administrative subdivisions of the states* are called municipalities (municipios), which are further subdivided into districts (distritos). For census purposes, the districts are subdivided into census enumeration areas (EAs).

Area (States)	Survey Area (km^2)	Area of Holdings (km^2)	Number of Holdings	Average Area of Holdings (ha)	Rural Population
Paraná	199 324	174 958	467 829	37.4	2 279 691
Santa Catarina	95 318	74 195	234 973	31.6	1 361 474
Distrito Federal	5 794	3 138	3 420	91.8	79 921
São Paulo	248 255	202 452	282 070	71.8	2 208 253
Total	548 691	454 743	988 292	46.0	5 929 339

Source: Brazil's 1985 Census of Agriculture and 1991 Population and Housing Census.

For area sample design purposes, it is useful to note that the *median* area of the holdings (establishments) is around 10 ha.

2. SURVEY OBJECTIVE AND SURVEY VARIABLES

The Brazilian General Purpose Annual Agricultural Survey Programme, based on multiple sampling frame methods, was established in order to obtain annual estimates of planted and harvested crop areas, areas intended for harvest, potential and actual yields, production for the major summer crops, livestock estimates, characteristics of the holdings and farming systems.

The main agricultural commodities studied are: apples, beans (first and second cropping), coffee, maize (first and second cropping), cotton, mangoes, manioc, onions, oranges, peanuts, potatoes (first and second cropping), rice (irrigated and non-irrigated), soybeans, sugar cane, tobacco and wheat; cattle, hogs, pigs and poultry.

The multiple frame survey design considers for each agricultural variable (in particular for each crop and livestock) the distribution with respect to the agricultural establishments of different size-groups. In fact, for many agricultural commodities, a high percentage of the total production is concentrated in a small number of large agricultural holdings. For instance, data from the 1985 Brazilian Agricultural Census shows that:

- Less than 1% of the agricultural holdings - those with more than 1 000 ha - account for around 44% of the total land.
- Apples: 7% of the holdings account for 69% of total production.
- Oranges: 10% of the holdings account for 57% of total production.
- Rice: 12% of the holdings account for 63% of total production.
- Soybeans: 7% of the holdings account for 63% of total production.
- Sugar cane: 9% of the holdings account for 81% of total production.
- Cattle: 2% of the agricultural holdings account for 33% of the total herd.
- Hogs and pigs: 10% of the holdings account for 20% of the national herd.

3. GENERAL CHARACTERISTICS OF THE SURVEY DESIGN

The general purpose Multiple Frame Agricultural Survey Programme is based on the application of large-scale probability sampling survey methods. The Multiple Frame Agricultural Survey Programme sample has two components: the *area sample component* which was selected once and used repeatedly over the years, and the complementary *list frame component* which is selected annually prior to the field data collection.

The annual field data collection, for both the area sample and the list sample components, is carried out by enumerators through direct interviews with the holders. Enlarged aerial photographs of the area sample segments are used to assist the field data collection, allowing the enumerator to obtain objective measurements of agricultural areas shown on aerial photography. The enumerators also make an inspection of all area sample segments.

The main characteristics of the multiple frame survey design are the following:

The area sample component of the multiple frame survey

The area sample survey design consists, in each *state*, of a one-stage, stratified, systematic and self-weighted probability sample of *segments*.

The area sampling frame considers the territory divided into a number of *strata* defined by the proportion of cultivated land, the predominance of certain crops or other land use characteristics. The strata and sample segments have *identifiable physical boundaries* (roads, paths, rivers, etc.) that can be located both in the field and on the cartographic materials used for their identification (satellite images from the TM sensor aboard the Landsat-V, mosaics of aerial photographs and maps, and field checks).

For the construction of the area sampling frame, the strata and administrative divisions of each state were completely subdivided into geographically ordered areas with recognizable physical boundaries called *counting units* (CUs) that, in some cases, were a subdivision or aggregation of contiguous *census enumeration areas*. The ordering of the counting units within a state, following a criterion of similarity, was done by ordering the municipalities within the state using cluster analysis techniques.

A systematic selection procedure of segments was applied and a sample of counting units and segments was systematically selected. In general, one segment within each selected counting unit was chosen. For the sample selection, each stratum was completely subdivided, according to similarity of agricultural characteristics or following geographic distribution criteria, into areas with an equal number of segments called *substrata* or *zones* (formed by entire CUs or parts of CUs), which provided a further level of stratification. In each substratum, segments were selected with equal probability of selection within the stratum. The area sample design can also be considered as a stratified, cluster sample of *tracts*, a tract consisting of the part of a holding (or non-agricultural areas) included in the segment.

For area frame construction, the delineation and transfer to different scales, and the area measurements of the administrative divisions, urban and rural areas, areas covered by water, strata and counting units were done by using an Intergraph computer graphic system.

The list sample component of the multiple frame survey

The survey design involved the construction and annual updating of complementary list frames of special agricultural holdings for each of the main crops and livestock herds surveyed. Each list is formed by holdings that correspond to a large percentage of the total area and production of the crop or to a large percentage of the total livestock herd.

Field data collection

The annual survey field data collection in the selected segments and special holdings list (the area sample and the list sample components) is done by enumerators through direct interviews with the holders. The enumerators complete a questionnaire for each agricultural holding partially or totally included in the sample segments or in each of the *special holdings*. The enumerators make an inspection of all area sample segments. An enlarged aerial photograph of each area sample segment is used to assist the field data collection, allowing the enumerator to obtain objective measurements of agricultural areas shown on aerial photography. The fields, holdings and other areas of specific land use within the selected segments are delineated on the aerial photographs covering the sampled segments. The measurement of areas within each segment is done in the field using a transparent grid, and later in the office using the computer graphic system.

Estimation methods

Weighted and closed multiple frame estimators, based on the combination of area sample data (data collected in the segments) and list frame data collected from the list of special holdings, were used to obtain the survey estimates and their coefficients of variation.

4. AREA FRAME CONSTRUCTION AND SAMPLE SELECTION

The general purpose annual Agricultural Survey Programme based on multiple sampling frame methods has been conducted in the State of Paraná since 1986. It was then extended to include the State of Santa Catarina and the Distrito Federal from 1987, and to the State of São Paulo from 1988. Since 1989, the Survey Programme has covered each year a very important Brazilian agricultural region, an area of 550 000 km^2.

4.1 Sample Size

The total area sample consists of 1 691 segments, distributed as follows: State of Paraná, 525 segments; State of Santa Catarina, 430 segments; State of São Paulo, 546 segments; and Distrito Federal, 190 segments.

4.2 Cartographic Materials Used for the Construction of the Sampling Frames and Sample Selection

- *Topographic charts*. Generally, charts in scales 1:50 000 and 1:100 000 were used. In the Distrito Federal, however, the charts used were in scale1:25 000 and dated 1985.
- *Satellite imagery* TM/Landsat-V on paper, in scales 1:100 000 and 1:250 000.
- *Index-map of TM/Landsat covering*.
- *Planimetric conjugates*, produced on a clear acetate, at the same scale as the TM/Landsat images, were used for visualization and geographical location of the interpreted patterns. The planimetric conjugate is the result of the superposition of the planimetric layer and the synonym of the topographic map. It originated from a map in scale 1:50 000, later reduced to 1:100 000.
- *Aerial photos* covering the selected segments. Aerial photos were available from companies in the original flight scale. Photo enlargements to 1:10 000 are used for the annual field data collection.
- *Photo indexes*. For example, the photo index of the Distrito Federal in scale 1:120 000, dated 1986.

- *Statistical municipal maps* (MMEs), prepared for census purposes. They show each municipality divided into enumeration areas. In order to improve the stratification process, the statistical agents were asked to mark in each MME the areas for main crops. Maps are generally in scale 1:100 000.
- *Political maps* and *Land use maps*.

4.3 Statistical Data for the Construction of the Sampling Frame

- 1985 Brazilian Agricultural Census (special tabulations). The Brazilian Censuses of Agriculture were conducted by full enumeration of all agricultural establishments, with five-year periodicity up to 1985. The latest census was conducted in 1996.
- 1985-1996 Agricultural Production Systematic Survey (LSPA). Monthly subjective survey at the state level.
- 1985-1996 Agricultural Production Municipal Survey (PAM). Annual subjective survey at the Municipal level.
- 1985-1996 Livestock Production Municipal Survey (PPM). Annual subjective survey that collects information at the Municipal level on the number of head of livestock, plus items on livestock products for the different herds.
- 1985-1996 Statistical Yearbooks. Published by IBGE annually.
- Special list frame agricultural surveys at the State level conducted by the State Secretaries of Agriculture - São Paulo, Paraná, Santa Catarina.
- Livestock vaccination records.

4.4 Land Use Strata Definitions, Target Segment Size and Sample Size in Each Land Use Stratum

The delineation of the strata was done by visual photo-interpretation techniques of satellite images supported by field inspection and also by using the cartographic materials available. Tables 1.1, 1.2, 1.3 and 1.4 show the definition of the strata adopted for the States of Paraná, Santa Catarina, Distrito Federal and São Paulo. In the tables, the following notation is considered:

L = number of land use strata in the state;
h = stratum subindex ($h = 1, 2, \ldots, L$);
i = substratum subindex ($i = 1, 2, \ldots, L_h$);
N = number of segments in the state;
N_h = number of segments in the h^{th} stratum;
N_{hi} = number of segments in the i^{th} substratum of the h^{th} stratum;
n = number of segments in the sample;
n_h = number of sample segments in the h^{th} stratum;
n_{hi} = number of sample segments in the i^{th} substratum of the h^{th} stratum;
$L_h = n_h / n_{hi} = N_h / N_{hi}$ = number of substrata in the h^{th} stratum;
$f_{hi} = n_{hi} / N_{hi}$ = sampling fraction of the i^{th} substratum in the h^{th} stratum.

4.5 Preliminary Delineation of the Strata

The first step was to delineate and measure the urban areas, rivers, lakes and other surfaces covered with water, and also the non-agricultural areas that were excluded from the stratification. The urban areas were those with more than 1 km^2. identifiable in the TM/Landsat-V satellite images in scale 1:100 000. For the remaining land, a preliminary visual photo-interpretation of the satellite images to identify the land use patterns was undertaken. At this stage, it was not intended to define precise strata boundaries.

State of Paraná. Panchromatic images, channels 3 and 4, scale 1:250 000 were used. Owing to the lack of colour enhancements in some areas and the implementation of special studies in a few areas of the state, a limited number of images, scale 1:100 000, channels 3 and 4 were also used. The preliminary stratification was done in transparent overlays on the images. Areas with less than 5 km^2 were not separated. In addition, each stratum had to cover at least 1% of the total area of the state. The information in the above-mentioned overlays was digitized using the Intergraph workstation and a cartographic base of the State of Paraná was also digitized and stored.

State of Santa Catarina. The TM images were also in the scale 1:250 000, but with a different colour composition, that is, channels 2, 3 and 4. As in the case of the state of Paraná, there was a need of ordering images in scale 1:100 000, channel 3 and/or 4, to obtain a better definition of the patterns.

Distrito Federal. An image, scale 1:100 000, colour composition, channels 2, 3 and 4, orbit 221/71, dated 25 February 1986, was used. For a detailed presentation of the preliminary stratification, see González-Villalobos et al. (1988).

State of São Paulo. Images in scale 1:100 000, colour composition, channels 3, 4 and 5, were used. This colour composition for São Paulo was considered to be the most suitable for stratification purposes.

4.6 Field Work for Improving the Strata Definitions and Boundaries. Definition of the Final Strata

The large-scale field inspection undertaken in the State of São Paulo for verifying the strata was of great importance for the stratification process. This procedure, incorporated in the construction of the frame, gave excellent results. The personnel in charge of the stratification travelled throughout the area, checking patterns of land use in all the strata. In the States of Paraná and Santa Catarina it was not possible to conduct as extensive a field checking as the one conducted in São Paulo. The field check made it possible to make improvements on the preliminary strata definitions by considering, for example, certain crop-specific strata. It also made possible the improvement of strata boundaries.

In the *State of Paraná,* a crop-specific stratum for potatoes was created. In the Distrito Federal, it was convenient to define special areas where the information would have to be collected using a different procedure. In these areas, removed from the original frame, the interviewers obtained the information by interviewing all operators, without using aerial photos.

In the *State of Santa Catarina*, the cultivation of apples is done in restricted areas, by small farmers. The availability of aerial photos made it possible to define the boundaries of a crop-specific stratum for apples used later in combination with a list of special farmers. Because of the economic importance of onion production, it was intended to define a crop-specific stratum for onions, but the idea was abandoned because of the characteristics of the cultivation done in very small scattered fields. This was confirmed by field trips.

In the *State of São Paulo*, a stratum of intensively cultivated land by small farmers was defined. This classification was introduced with the purpose of reducing the field workload.

4.7 Delineation of Final Strata Boundaries. Transposition to the Planimetric Conjugates

The strata boundaries were transferred from the satellite images to topographic maps in scale 1:100 000, and the strata boundaries were adjusted to identifiable permanent physical limits. In the States of Paraná and Santa Catarina, as already mentioned, the preliminary strata boundaries were delineated in transparent acetate overlays on the satellite images. Then, the preliminary strata boundaries were transferred to transparent acetates in scale 1:100 000, generated by the Intergraph workstation. These transparent overlays were used on the maps to adjust the strata boundaries to follow identifiable permanent physical boundaries.

State of Paraná. Using a special set of topographic charts in scale 1:250 000, city boundaries, rivers, lakes, roads and other permanent physical boundaries were also digitized. Then the boundaries of the 311 municipalities were stored. The Intergraph subsequently generated 93 transparent overlays, in scale 1:100 000, which included the cartographic base, the boundaries of all municipalities and the preliminary strata. The overlays were used on light tables, jointly with the available satellite images in the same scale and with the topographic maps in scales 1:100 000 and 1:50 000, to adjust the strata boundaries to permanent physical boundaries.

Distrito Federal. The above operations were simplified because the satellite imagery used was in the scale 1:100 000. For a detailed presentation of the final operations for stratification see González-Villalobos, et al. (1988), section 6.4.

State of São Paulo. The operations were improved since the satellite images used were all in the scale 1:100 000, and the acetates were substituted by planimetric conjugates. The planimetric conjugates were important for the construction of the frame since the different boundaries, and in particular the strata boundaries, were delineated directly on them.

4.8 Measuring the Area of the Strata and Allocation of the Sample

The calculation of the area of the strata was done by using the Intergraph workstation. The sample allocation to the strata was determined as a function of the strata measures of size, the internal homogeneity in terms of number of segments and the number and type of the main variables under consideration.

4.9 Counting Units: Definition, Delineation, Measurement and Assignment of Size Measures

Construction of Counting Units

The strata were subdivided into measured areas called counting units (CUs) and indicated in the maps. The CUs were formed by delineating continuous land areas, with well-defined physical boundaries (roads, paths, rivers, etc.) within each stratum and administrative division, and containing an approximately equal number of segments. Where possible, the CUs were formed by subdivisions or aggregation of contiguous 1985 Agricultural Census EAs.

In the States of Paraná, Santa Catarina and São Paulo, the following criteria for the determination of the counting unit (CU) target size in each stratum were applied:

Assuming that all CUs in the stratum have approximately equal size (number of segments), the target size was obtained by minimizing the function f(x):

$$f(x) = (N_h / \bar{x}_h) + \bar{x}_h \cdot n_h$$

which considers the effort to subdivide the strata into CUs and the CUs into segments.

If the derivative of f(x) is equal to 0, then

$$N_h / \bar{x}_h^2 = n_h$$

Therefore, the target segment size of the CUs in the h^{th} stratum would be

$$\bar{x}_h = \sqrt{N_h / n_h} = \sqrt{1 / f_h}$$

State of Paraná. A total of 6 926 CUs were defined. This work was done by 30 technicians during two months.

Distrito Federal. In most cases, the boundaries of the CUs followed those of the 43 census enumeration areas and 209 subenumeration areas. If T denotes the target area of a segment in a given stratum, then the area of each CU in the stratum was such that:

- In most cases approximately equal to 4T.
- Never larger than the inverse of the sampling fraction times T.

Measuring the area of counting units and assignment of size measures

The calculation of the CU areas was also done by using the Intergraph workstation. Measures of size were assigned to the CUs. The measure of size of a CU was established to be equal to the closest integer to the quotient between its area and the area of the target segment in the stratum.

4.10 Ordering the CUs. Geographical Substratification

Geographical substratification is a second level of stratification which was applied in order to improve the efficiency of the design. The number of segments by substratum had to be a multiple of five. This procedure would allow the annual rotation of 20% of the sample segments reducing the respondent burden caused by repeated interviewing.

Within each land use stratum, the geographical substrata were formed by entire CUs, or parts of CUs, with similar agricultural characteristics and an equal number of segments except for the last CU of each stratum which could have a slightly different number of segments.

In the process of constructing such a frame, the municipalities were ordered within the State using techniques of cluster analysis, with a restriction of contiguity. To reduce the dimension of the cluster analysis, principal components analysis and factorial analysis were applied. Techniques of hierarchical clustering were applied to the municipalities producing the tree used in the ordering process. The selected variables considered in the hierarchization process, related to the survey's variables, were the following: land use; farming structure; technological level; livestock; and agricultural products investigated in the survey. With this ordering it was possible to create the geographical substrata.

4.11 Sample Selection

In each substratum n_{hi} segments were systematically selected, with equal probability and without replacement (EPSEM). The determination of the number of sample segments per CU and the selected CUs in each substratum was accomplished by using the accumulated measures of size of the CUs, following the established order. Each selected CU was subdivided into a number of segments equal to its measure of size. The selected segment/segments in each CU were those indicated by the systematic selection.

The n_h values, resulting from the allocation of the sample to the strata, were adjusted to multiples of five. This selection scheme was used in the samples of Paraná, Santa Catarina and São Paulo. In the Distrito Federal the sample design did not include substrata.

4.12 Preparation of the Aerial Photographs of the Selected Segments

Each selected segment was identified and delineated, first on the conjugates or on the acetates and then on the photo index, in order to identify the corresponding aerial photos. With the photos identified, copies were ordered from the aerophoto companies, in the flight scale, and an enlargement in the scale of 1:10 000 for data collection purposes. In some cases, depending on the size of the segment and the precision of the identification, it was necessary to order more than one photo to cover the whole area of the segment. When the photos were received, in both scales, the segment boundaries were delineated: first, directly on the photo which is in the flight scale; second, on an acetate which covers the enlargement, which is the basic material for the enumerators' work. Then the material was sent to the State Statistical Offices.

5. THE COMPLEMENTARY LIST SAMPLE FRAMES OF SPECIAL HOLDINGS

Based on the 1985 Agricultural Census information, several lists of holdings that concentrate a large percentage of the total of the variable were constructed. These lists, including a relatively small number of holdings, are called Lists of Special Holdings and are updated every year.

The following survey variables and list frames of special holdings were used for the multiple frame estimators in the 1994 Agricultural Survey:

List frames in the State of Paraná

Cotton: 32 agricultural holdings, with an area ≥100 ha, selected from a frame of 150 holdings, corresponding to an area of 29 750 ha.
English potatoes: 80 holdings, totally included in the investigation, corresponding to an area of 2 936 ha.
Sugarcane: 41 holdings, with an area ≥200 ha, selected from a frame of 139 holdings, corresponding to 106 525 ha.
Cattle: 259 holdings with more than 2 000 head, including 884 708 head or approximately 40% of the state herd.

List frames in the State of Santa Catarina

All agricultural holdings of the list frames were canvassed.

Apples: 11 agricultural holdings, corresponding to 6 292 ha.
Sugar cane: 1 holding, corresponding to 4 477 ha.
Hogs and pigs: 72 agricultural holdings, corresponding to 335 284 head.

List frames in the Distrito Federal

All the agricultural holdings of the list frames were canvassed. The list frames in the special strata include, in particular, all holdings with 500 ha or more.

Beans: 9 agricultural holdings, with an area ≥50 ha.
Coffee: 8 agricultural holdings, with an area ≥50 ha.
Maize: 11 agricultural holdings, with an area ≥200 ha.
Rice: 9 agricultural holdings, with an area ≥50 ha.
Soybeans: 26 agricultural holdings, with an area ≥300 ha.

Cattle: 14 agricultural holdings, with ≥500 head.
Hogs and pigs: 3 agricultural holdings, with ≥200 head.

List frames in the State of São Paulo

The following variables and list frames of special holdings were used for the multiple frame estimators:

Cotton: 72 agricultural holdings, with an area ≥50 ha, selected from a frame of 598 holdings, corresponding to an area of 61 337 ha.
Oranges: 95 agricultural holdings, with an area ≥100 ha, selected from a frame of 781 holdings, corresponding to an area of 247 132 ha.
Cattle: 35 agricultural holdings, with ≥3 000 head, selected from a frame of 181 holdings, corresponding to 940 706 head.
Hogs and pigs: 26 agricultural holdings, with more than 1 000 head, selected from a frame of 93 holdings, corresponding to 221 301 head.

6. ESTIMATION PROCEDURES. SAMPLING ERRORS

For variables with a distribution such that a large percentage of the total is concentrated in a relatively small number of the holdings, multiple frame estimators were used in order to improve the precision of the direct area sample estimators.

For a given variable, the multiple frame estimator is the sum of the estimators from both samples, the area sampling estimator and the list sample estimator based on the list frame of special holdings.

6.1 Area Sample Estimators

Two direct area sampling frame estimators were used to estimate the state totals for a particular variable: the *weighted segment estimator* or the *closed segment estimator*. The choice of the type of estimator for each variable was based on the corresponding reporting unit (tracts or holdings).

The formulae used for the direct expansion area sample estimators and their variances are presented in Chapter 11.

6.2 List Sample Estimators

The holdings included in the list of special holdings (E) were excluded from the selected segments of the area frame in order to avoid duplication of information. In the case that E included a large number of holdings, or when the field data collection was expected to be difficult to accomplish, a stratified simple random sample of holdings was selected from E. If E included a small number of holdings or if the field data collection was easy to accomplish, all the holdings in E were investigated.

6.3 Multiple Frame Estimators

The corresponding multiple frame survey estimators were used for the estimation of totals and coefficients of variation (CVs) which are the square root of the relative variance of the sample estimates and determine the range of the confidence intervals for the estimates.

The formulae used for the direct expansion dual frame estimators and their variances are presented in Chapter 11.

7. DATA COLLECTION

Data collection period

The crop calendar showed the months of planting and harvesting of the crops surveyed, according to the states where the surveys have been conducted. The last quarter of the year was chosen in order to have an early estimate of the planted areas for the summer crops, before harvesting. The annual survey field data collection is conducted during the period October to December, and lasts approximately 40 days.

It is readily seen that some crops do not follow a perfectly defined cycle. Thus the date of the survey field data collection period was established in order to avoid losing a considerable amount of reported planted area. The first rounds of the survey were conducted during one month in the first quarter of the year; for example, in the State of Paraná the first survey was conducted from 19 January to 24 February 1987 - a period of 35 days. Later on, with the experience gained in conducting several survey rounds, the period was changed since it was found that the heavy rains during such a period contributed to increase the survey's costs and resulted in delays in the field work.

Selection and training of field personnel

Since data collection experience and good knowledge of the survey area were considered essential conditions for the selection of field personnel, the data collection personnel were largely selected from Brazilian National Statistical Office regional offices. Enumerators were trained for a few days prior to the survey and received additional supervision while in the field. An Enumerator's Manual was prepared. In each state, a five-day theoretical training course was given followed by an examination. A five-day practical training course was also given. A driver with a car was assigned to each group of enumerators. The first part of the training programme concerns the theoretical aspects of the survey fieldwork: the technical knowledge required, general survey characteristics and the survey concepts and definitions used during the data collection. The second part consists of the practical training in data collection directly in the field. In this part the enumerators, divided into groups of four, go through all of the stages of the data collection. The end of the training course is dedicated to the supervisors who receive special instructions for their work.

The field data collection of each survey round is undertaken by a total of 20 supervisors and 200 enumerators.

Coordination of the field data collection

In each state, a survey coordination unit was created, composed of one coordinator, the supervisors and the enumerators. IBGE maintains an extensive network of field offices in each of the surveyed states. These State Statistical Offices provide the infrastructure for the survey, the logistics for sending and receiving survey materials, vehicles, communications, etc.

Field data collection materials

- *Enumerator's manual.*
- *Municipal statistical map and a cartographic map* (and/or state or county highway maps) showing the segment boundaries. Included were detailed written instructions for arriving at the segment from the closest village.
- *Listing sheet.* This form is used to list all agricultural holdings totally or partially included in the segment. The listing sheet is used for the identification of the agricultural holdings for which a questionnaire should be completed.
- *Questionnaire.* Each questionnaire corresponds to one agricultural holding totally or partially included in the segment, as reported in the listing sheet. The questionnaire covers, in particular, the following subjects: general characteristics of the agricultural holding; planted area, area to be planted and harvested area, average yield, expected and actual yields, planting and harvesting dates for each crop; characteristics of the cattle, hog and pig herds; amount of seed, by type.
- *Aerial photos*, scales 1:10 000 and 1:25 000, with the delineation of the segment, on a transparent overlay. The overlays on the aerial photos are used to delineate areas during the field data collection.
- *Miscellaneous materials.* Ruler, eraser and blue, red and green port-crayons used to delineate different types of areas in the photo overlays.
- *Grid of points*, used to measure areas on the aerial photos during data collection.

Field data collection procedures

In each selected segment the team of enumerators performs the following activities :

- Identification in the field of the segment area.
- Completion of the *listing sheet*, listing all agricultural establishments totally or partially included in the segment.
- Delineation on the transparent overlay, which covers the aerial photo in scale 1:10 000, of all holdings totally or partially included in the segment.
- Completing a questionnaire for each one of the establishments totally or partially included in the segment. The questionnaire is completed by interviewing the operators. The enumerator is responsible for obtaining the information about the planted crops, the planting intentions and on the uses of the land for each particular field. The enumerator is also responsible to obtain the information related to the other items of the questionnaire.
- Drawing all fields on the above-mentioned aerial photo overlay, in order to identify the use of the land in each field of the segment.

- Measuring each field delineated in the overlay, using the grid of points, in order to check the operator's answers.

8. DATA PROCESSING

The IBGE state local offices check the survey material as soon as the data collection is completed. The material is then sent to the IBGE State Office for data entry and transmittal to IBGE headquarters in Rio de Janeiro. An IBM 3090 computer, DB2 language and SAS software are used. The survey data are processed as follows:

- Manual data editing.
- Verification of the consistency of the data contained in the listing sheets and questionnaires.
- Verification and new measurement of areas included in the overlays on the aerial photographs. Comparison with the data presented in the listing sheets and questionnaires.
- Computer data editing. Application of special programs for quality control and imputation of missing or defective data.
- Tabulation and analysis of non-weighted survey data.
- Weighting of survey data. Calculation of the different types of direct area sampling frame estimates and multiple frame estimates.
- Tabulation of the final results including presentation of coefficients of variation.

9. SURVEY DATA ANALYSIS

The main estimates of each annual round of the Multiple Frame Survey in each state were compared with the estimates obtained from the most important independent sources, namely: the 1985 and 1996 Brazilian Agricultural Censuses, conducted by full enumeration of all agricultural holdings; the Monthly Agricultural Production Systematic Survey (LSPA), a subjective survey at the state level; and the Annual Livestock Production Municipal Survey (PPM), a subjective survey at the municipal level.

The Agricultural Production Systematic Survey (LSPA) is a subjective survey, whose estimates for 56 commodities are derived from regular meetings, conducted by IBGE State Offices with the participation of specialists of several organizations from the agricultural sector. These estimates, corresponding to crops cultivated by small farmers without official financing or technical assistance, are likely to be less accurate, as in the case of rice, beans and manioc. On the other hand, the estimates corresponding to crops with official financing or technical assistance are likely to be more accurate, as in the case of cotton, soybeans, maize, oranges or apples, since there are several sources for checking and controlling the estimates. For these reasons, some LSPA crop forecast estimates and PPM estimates use Multiple Frame Survey results.

The survey data analysis involved, for each survey variable, the study of coefficient of variation for the different survey rounds, the treatment of outliers as well as the above-mentioned comparison of results with alternative sources.

10. GEOGRAPHIC INFORMATION SYSTEM FOR THE SURVEY PROGRAMME

The area frame construction involved the preparation of overlays and planimetric conjugates, generated by the Intergraph and Maxi-Cad workstations, showing the boundaries of the areas involved in the different steps: strata and CUs. In 1995, the sampled segments were also digitized and, therefore, a Geographic Information System (GIS) for the Survey Programme was prepared, indicating all steps of the area frame construction in each state.

The GIS has improved the capability of analysing annual changes in the segments, as well as many other advantages.

The segment delineation was georeferenced in the topographic chart by using the middle point of the smallest square that included the segment. This was generated by ARC/INFO files and exported for visualization in PC/Windows through ARC/VIEW 1.0 software.

The visualization combines layers showing municipalities, districts, strata, counting units, segments and tracts; and obviously the corresponding data of each survey round.

The GIS makes it possible to combine the information of this large database for the different needs of data users.

11. FUTURE DEVELOPMENTS OF THE MULTIPLE FRAME SURVEY PROGRAMME

- Extension of the survey area to cover several new states: Mato Grosso do Sul and Rio Grande do Sul.
- Studies on area sampling design improvements to carry out that extension.
- New updated area and list frames should be constructed. For this purpose, the 1996 Agricultural Census data and cartography should be used. For the construction of the new area frames, new target segment sizes should be considered.
- Implementation of the already prepared Objective Wheat Yield Survey Programme, based on the area frame.
- Implementation of the already prepared Environmental and Land Use Survey Programmes, based on the area frame.
- More data analysis is needed and more detailed technical and non-technical documentation is required.

Table 1.1
Brazil: area sampling frame design information
State of Paraná - 1991

Strata definition	Survey area (km^2)	Target segment size (km^2)	Number of CUs	Total segments N_h	Sample segments n_h	Segments in substrata n_{hi}	Number of substrata L_h	Size of substrata N_{hi}	$1/f_h$	Counting unit size: Maximum – Number of Segments	Counting unit size: Maximum – Area (km^2)	Counting unit size: Target – Number of Segments	Counting unit size: Target – Area (km^2)
Intensive agriculture, 80% and more of cultivated area	29 814	1	1 774	29 808	135	5	27	1 104	220.8	110	110	15	15
Intensive agriculture, 50% to 79% of cultivated area	19 062	1	1 158	19 120	85	5	17	1 125	225.0	110	110	15	15
Special stratum - potatoes	2 194	2	81	1 095	30	15	2	547	36.5	20	40	6	12
Extensive agriculture - between 15% and 49% of cultivated area, with pastureland	23 915	2	801	11 916	55	5	11	1 083	216.6	60	120	11	22
Extensive agriculture - between 15% and 49% of cultivated area, without pastureland	5 0 307	2	1 809	25 151	165	5	33	762	152.4	60	120	11	22
Extensive agriculture - less than 15% of cultivated area	29 963	4	548	7 513	35	5	7	1 073	214.6	110	440	15	60
Non-agricultural land	39 512	8	755	4 937	20	5	4	1 234	246.8	130	1 040	16	128
Total	**194 797**	-	**6 926**	**99 540**	**525**	-	-	-	**189.6**	-	-	-	-

NOTE: Total land area of the state equals 199 323.90 km^2. Sample segments correspond to 0.5% of the total land area of the state.

Table 1.2
Brazil: area sampling frame design information
State of Santa Catarina - 1991

Strata definition	Survey area (km^2)	Target segment size (km^2)	Number of CUs	Total segments N_h	Sample segments n_h	Segments in substrata n_{hi}	Number of substrata L_h	Size of substrata N_{hi}	$1/f_h$	Counting unit size: Maximum: Number of Segments	Counting unit size: Maximum: Area (km^2)	Counting unit size: Target: Number of Segments	Counting unit size: Target: Area (km^2)
Intensive agriculture, 80% and more of cultivated area	2 010	1	289	2 010	6	1	6	335	335.0	20	20	6	6
Intensive agriculture, 50% to 79% of cultivated area	4 981	2	318	2 491	30	5	6	415	83.0	50	100	9	18
Special stratum - apples	80	1	15	80	10	5	2	40	8.0	5	5	3	3
Special stratum - several products	2 483	1	427	2 483	24	24	1	113	103.4	10	10	5	5
Extensive agriculture - between 30% and 49% of cultivated area	2 686	2	907	11 344	210	15	14	810	54.0	50	200	9	36
Extensive agriculture - between 15% and 29% of cultivated area	24 602	2	945	12 302	75	5	15	820	164.0	50	200	9	36
Extensive agriculture - less than 15% of cultivated area	9 394	4	293	2 348	30	5	6	391	78.2	50	200	9	36
Rangeland with less than 10% of cultivated area	17 862	8	394	2 233	30	5	6	372	74.4	40	320	9	72
Non-agricultural land	9 918	8	209	1 240	15	5	3	413	82.6	40	320	9	72
Total	**94 016**	-	**3 797**	**36 531**	**430**	-	-	-	**84.96**	-	-	-	-

NOTE: Total land area of the state equals 95 318.3 km^2. Sample segments correspond to 1.2% of the total land area of the state.

Table 1.3
Brazil: area sampling frame design information
Distrito Federal - 1991

Strata definition	Survey area (km^2)	Target segment size (km^2)	Number of CUs	Total segments N_h	Sample segments n_h	$1/f_h$	Counting unit size: Maximum: Number of Segments	Counting unit size: Maximum: Area (km^2)	Counting unit size: Target: Number of Segments	Counting unit size: Target: Area (km^2)
Intensive agriculture, 60% and more of cultivated area except mangoes	536	1	101	536	33	16.24	16	16	4	4
Intensive agriculture, 40% to 59% of cultivated area	381	1	94	381	23	16.57	16	16	4	4
Extensive agriculture - between 20% and 39% of cultivated area	1 337	1	191	1 336	79	16.91	16	16	4	4
Extensive agriculture - up to 20% of cultivated area without horticulture	1 746	3	189	583	35	16.65	16	36	4	12
Extensive agriculture - up to 20% of cultivated area with some horticulture	79	3	8	26	2	13.00	13	36	3	9
Extensive agriculture - up to 20% of cultivated area with horticulture	228	1	50	226	14	16.14	16	16	4	4
Silviculture	224	3	19	75	4	18.75	18	49	5	15
Special stratum - mangoes	9	*	*	*	*	*	*	*	*	*
Other agricultural area	184	*	*	*	*	*	*	*	*	*
Non-agricultural land	1 070	*	*	*	*	*	*	*	*	*
Total	**5 794**	-	**652**	**3 163**	**190**	**16.64**	-	-	-	-

NOTE: Total land area of the state equals 5 794.2 km^2. Sample segments correspond to 4.7% of the total land area.

Table 1.4
Brazil: area sampling frame design information
State of São Paulo - 1991

Strata definition	Survey area (km^2)	Target segment size (km^2)	Number of CUs	Total segments N_h	Sample segments n_h	Segments in substrata n_{hi}	Number of substrata L_h	Size of substrata N_{hi}	$1/f_h$	Counting unit size			
										Maximum		Target	
										Number of Segments	Area (km^2)	Number of Segments	Area (km^2)
Intensive agriculture, 80% and more of cultivated area: small fields	5 116	0.5	816	10 237	36	3	12	853	284.36	17	10	95	45
Intensive agriculture, 80% and more of cultivated area: large fields	17 016	1	1 140	17 027	70	3(*)	21	811	243.24	19	20	95	95
Sugar cane	22 811	2	979	11 393	35	5	7	1 628	325.51	18	35	100	200
Citrus	2 865	2	143	1 439	5	5	1	1 439	287.80	17	35	140	250
Between 50% and 80% of cultivated area, plus pastures and/or woodland: medium and large fields	15 807	2	670	7 896	65	5	13	607	121.48	11	25	95	200
Between 50% and 80% of cultivated area, plus pastures and/or woodland: small fields	14 497	0.5	2 430	28 991	129	3	43	674	134.84	12	5	95	45
Between 30% and 50% of cultivated area, plus pastures between 30% and 50% (pastures may be together with woodland): medium and large fields	25 710	3	869	8 563	65	5	13	659	131.74	11	35	100	300
Between 30% and 50% of cultivated area, plus pastures between 30% and 50% (pastures may be together with wóodland): small fields	8 601	1	553	8 598	69	3	13	661	132.28	12	10	95	100

(continued)

Table 1.4 (cont'd)
Brazil: area sampling frame design information
State of São Paulo - 1991

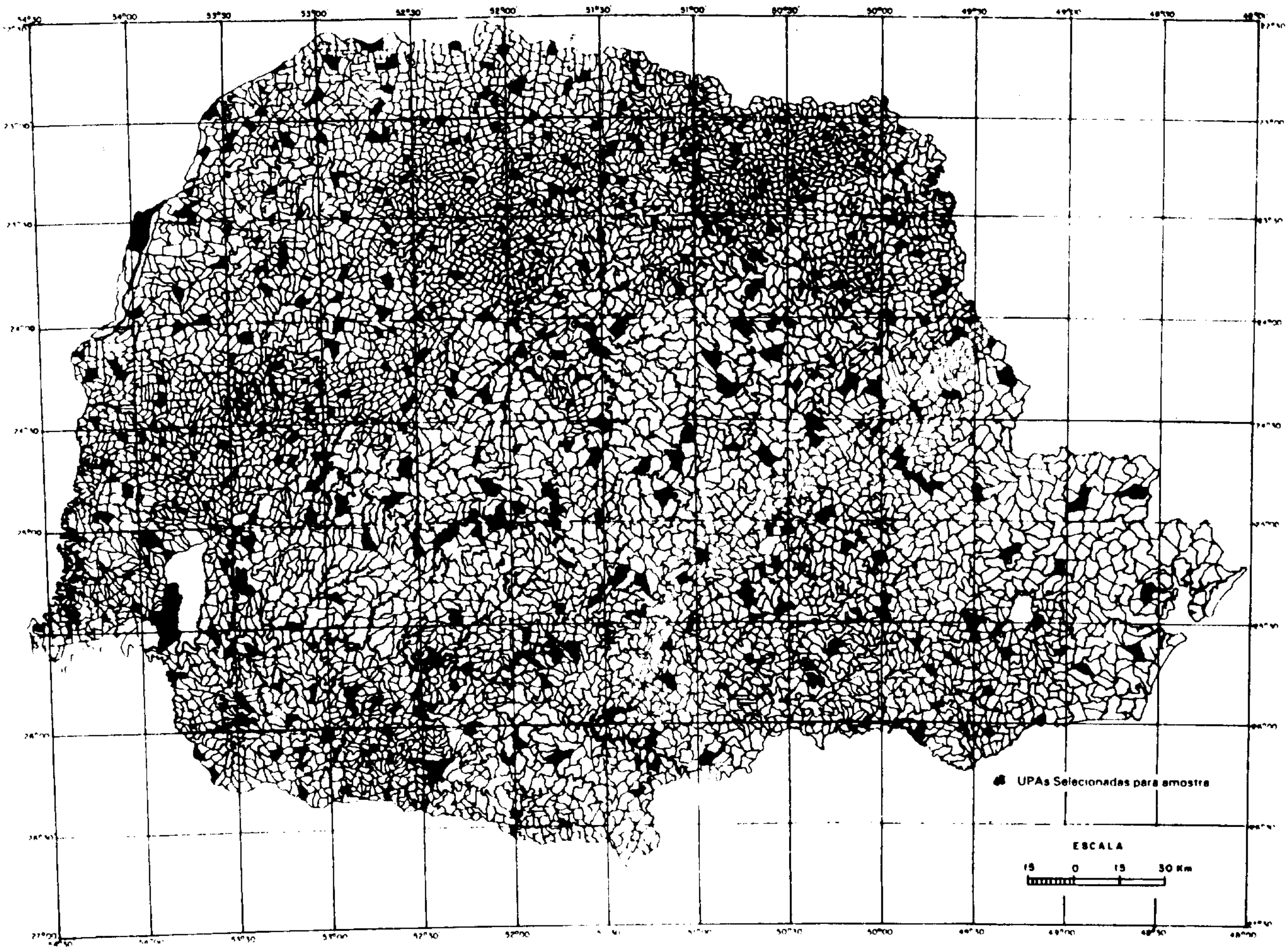
Strata definition	Survey area (km^2)	Target segment size (km^2)	Number of CUs	Total segments N_h	Sample segments n_h	Segments in substrata n_{hi}	Number of substrata L_h	Size of substrata N_{hi}	$1/f_h$	Counting unit size			
										Maximum		Target	
										Number of Segments	Area (km^2)	Number of Segments	Area (km^2)
80% and more of pastures: medium and large fields	32 330	4	855	8 088	30	5	6	1 348	269.60	16	65	100	400
Between 50% and 80% of pastures, plus agriculture and/or woodland: medium and large fields	39 054	4	1 047	9 772	30	5	6	1 629	325.73	18	70	100	400
Between 50% and 80% of pastures, plus agriculture and/or woodland: small fields	6 169	2	561	3 089	10	5	2	1 544	308.90	18	35	100	200
Between 50% and 80% of pastures, plus woodland: medium and large fields	9 869	4	373	2 457	5	5	1	2 457	491.40	22	90	95	350
80% and more of woodland	27 856	4	811	6 964	20	5	4	1 741	348.20	19	75	100	400
Between 50% and 80% of woodland, plus agriculture and/or pastures	10 032	4	364	2 502	5	5	1	2 502	500.40	22	90	80	350
Non-agricultural land	154	1	28	154	2	2	1	154	77.00	9	10	40	40
Total	**237 887**	-	**11 639**	**127 170**	**546**	-	**145**	-	**193.27**	-	-	-	-

NOTE: Total land area of the state equals 248 255.7 km^2. Sample segments correspond to 0.4% of the total land area of the state.

(*) The fifth substratum includes ten segments.

Figure 1.1

**The State of Paraná divided into administrative divisions and PSUs.
Selected PSUs shown in black
(figure obtained by using the Intergraph Workstation)**



Source: Revista Brasileira de Estatística, No. 191, p. 63, 1988.

State of Santa Catarina: Land use stratification

(obtained by using the GIS)

Figure 1.2

Overlay of a segment aerial photo indicating holding and tract boundaries

(obtained by using the GIS)

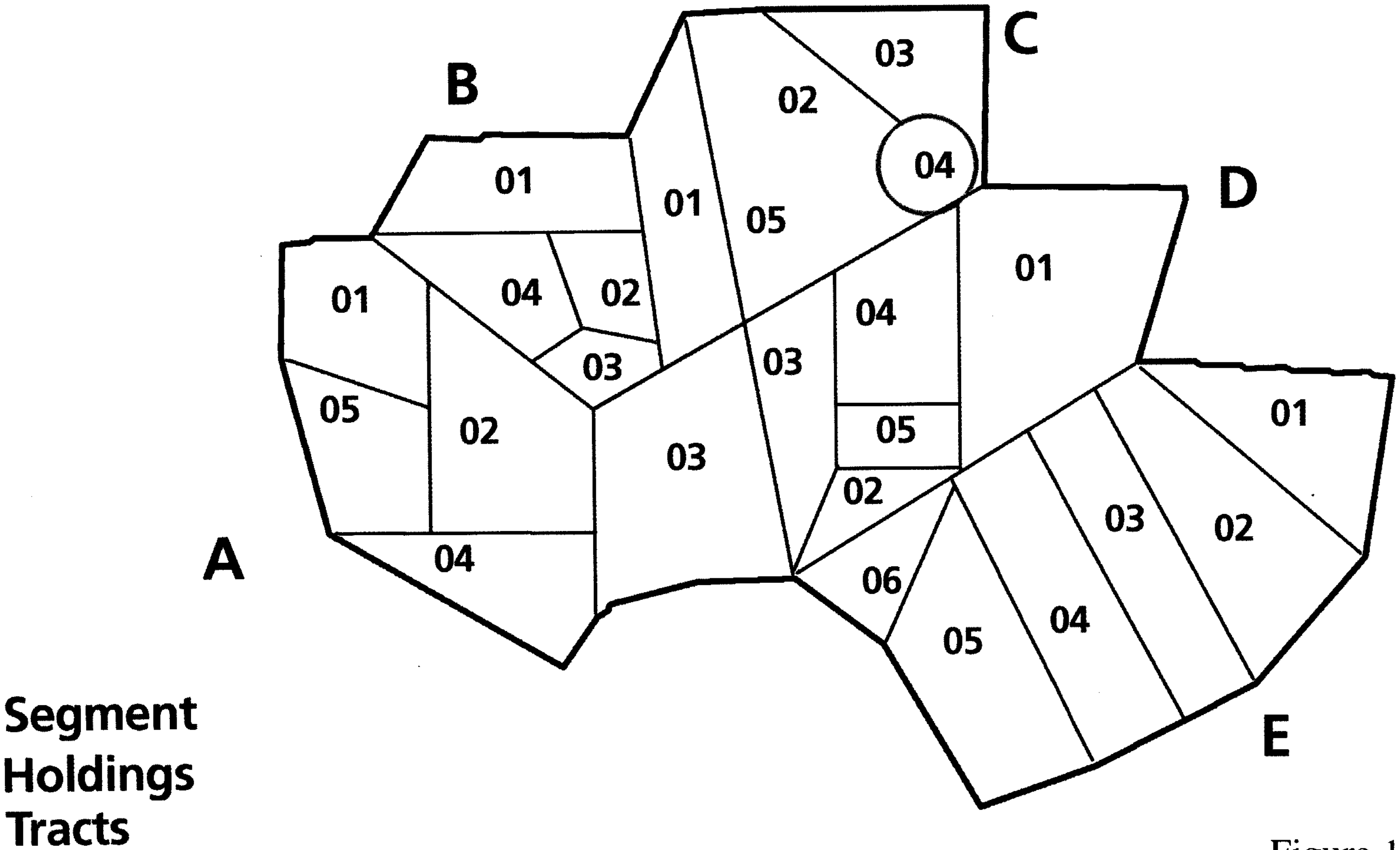


Figure 1.3

CHAPTER 2

CANADA

GENERAL PURPOSE AGRICULTURAL SURVEY PROGRAMME BASED ON MULTIPLE FRAME SAMPLING METHODS (1988-1995)

1. INTRODUCTION

This chapter describes Statistics Canada's experience with implementing a general purpose multiple frame survey programme to collect agricultural statistics in Canada. The focus of the case study is on operational considerations and the reasons why this approach was chosen.

Since the 1950s, area frame survey designs have been used for the regular agricultural surveys in the intercensal years. The area frames for the Canadian agricultural surveys were based on maps and data from the Census of Agriculture.

It was not until the late 1970s that the use of more than one type of frame was considered for Statistics Canada's agricultural surveys. This move towards a multiple frame survey programme was driven by the need to increase the efficiency of the design while controlling costs and ensuring up-to-date coverage of the population.

From 1983 to 1992, a multiple frame sample design was used for the annual National Farm Survey. Since 1993, following the 1991 Census of Agriculture, the agricultural area sample survey was redesigned and was called Area Farm Survey (AFS). The AFS is used to complement the list samples of various other agricultural surveys with sometimes different list frames.

List sample designs* vis-a-vis *area sample designs. As explained in Volume I, Chapter 2, large list frames have the problem of incompleteness that increases with age. Furthermore, constructing list frames by merging lists from different sources is costly and can include undetected duplicates. If the list frame has good addresses, data can be collected by mail. However, mail surveys tend to have low response rates, demanding repeated follow-ups. Usually data from mail surveys need to be edited more carefully than data collected through personal interviews, thus raising costs.

With increasing telephone dissemination, more and more list frames surveys are moving towards telephone interviews, which are less expensive than personal interviews. Despite their higher cost compared with mail surveys, telephone surveys tend to have better response rates and allow the data to be edited while the respondent is on the telephone.

Theoretically, area frames provide complete coverage of the population if the collection of geographical areas or segments that comprise the frame covers the entire population of interest. An advantage of an area frame is that the population is clearly defined, and all agricultural operations have the potential to be identified. An area frame is a more durable representation of the population of interest compared with a list frame which can become quickly out of date. From a fieldwork point of view, area "segments" are clearly identifiable on maps and in the field, provided the boundaries are explicitly and correctly marked; entries on a list are more conceptual in nature.

However, the effectiveness of an area frame depends on the availability of good maps, size measures and personnel experienced in field enumeration. The creation of an area frame usually requires more resources than the development of a list frame.

Multiple frame sample design. In general, a multiple frame design requires a set of frames that together cover all the units in the population. It is essential that every unit in the population of interest is contained in at least one of the frames, and that it is possible to identify all the frames that contain each sampled unit. In most applications, a two-frame design is used with one frame that covers all units in the population, while the other frame covers only a fraction of the population, usually the largest units (Hartley, 1962).

To implement a multiple frame design in Canada, an area frame was the only option for the "complete" frame since an up-to-date list frame that covers the population in its entirety does not exist. A multiple frame approach also allows different sample designs to be implemented for each of the frames. This allows the sample from the list frame to be an efficient unclustered single-stage sample that can also focus on certain aspects such as providing small area estimates or estimating rare populations, while the sample from the area frame can focus on compensating for the undercoverage of the list frame. Among others, two characteristics of Canada's agricultural sector have influenced the multiple frame designs adopted, namely: the wide availability of telephones in the rural area, and the five-year frequency of the complete enumeration Agricultural Census undertaken jointly with the Population and Housing Census.

Implementation of a multiple frame design. Statistics Canada has access to the Census of Agriculture data that serve as a "single" list frame whose coverage is about 97 to 98% complete in the census year. There is no need to merge and unduplicate many lists to create the list frame. However, the Census of Agriculture is only conducted every five years and census data can quickly become out of date as the agricultural sector changes. New farms originating after the last census can be covered using an area frame to supplement the list frame. Since the Census of Agriculture in Canada is conducted at the same time as the Census of Population and uses the same enumerators, the enumeration areas (EAs) created for the Population Census serve as useful units for the area frame.

The detection of the overlap between the list frame and the area sample is one of the most crucial steps in the multiple frame survey process. The automation of the overlap detection process through the use of computerized matching or probabilistic record linkage techniques could improve the quality of this important operation.

The cost of constructing and maintaining the area frame is higher than that of the list frame; however, it is felt that these costs are justified because the identification of new farms is sufficiently important.

2. AGRICULTURAL CHARACTERISTICS, SURVEY AREA AND SURVEY VARIABLES

Canada covers 10 million km^2 and spans 5 500 km from east to west; almost 1 million km^2 is cultivable land. Politically, the country is divided into ten provinces and two territories.

The agricultural sector is the second largest primary industry in the country, contributing 10% to its gross domestic product. Over one-half of the agricultural products produced in Canada are exported. In 1991, Canada had 280 183 census farms operated by 390 870 individuals. These farms cover 68 million ha of land, of which 45 million ha are cultivated. The number of farms in Canada has been decreasing continuously since 1941, resulting in fewer but larger farms. According to the 1991 Census of Agriculture, 98% of these farms were family operated and only 1% were operated by a corporation. The remainder were institutional farms or farms on Indian reserves (cf. Table 2.1).

Estimates are often required by farm type, which is determined by the predominant commodity of the farming operation. There are four main types of farms in Canada: *livestock farms* are specialized in beef production and cattle, hogs, poultry, sheep and dairy cattle raising; *grain and oilseed farms* primarily produce wheat, canola, barley, rye, maize, soybeans, oats and flax; *mixed farms* produce both grains and livestock; *special crop farms* primarily produce vegetables, fruits, root crops, tobacco, greenhouse and nursery products or forest products.

The Canadian agricultural surveys cover all provinces except the Yukon and the Northwest Territories which have little agricultural activity. Newfoundland usually has a survey specifically designed for that province since its agricultural activities are different from those of the other provinces and they contribute very little to national level estimates.

3. DESIGN OF THE NATIONAL FARM SURVEY (1988)

The National Farm Survey (NFS), conducted from 1983 to 1992, was an annual, general purpose, probability-based sample survey which collected data on crops, livestock and farm finances.The objective of the survey was to provide timely, reliable estimates for major variables at the provincial level with a target coefficient of variation (CV) of 5%. However, estimates were also required at subprovincial levels. The target population of the NFS was all farms which received or expected to receive at least $Can 250 annually from the sale of agricultural products. The survey population was covered by a list frame and an area frame, both of which were based on the 1986 Census of Agriculture.

Farms on Indian reserves and institutional farms were excluded from the survey population because it is difficult to collect data from them. The census estimated their contribution to the estimates at the national level to be 0.1%. This estimate was used for adjusting the final survey estimates.

The use of a multiple frame design with both a list and an area component for the NFS provided precise and timely estimates for ten years that met the targeted coefficient of variation of 5% at the Canada level for major variables. The annual contribution of the area frame component to the estimates, which accounted for census undercoverage and new farming operations, was 4% to 10% across Canada.

Data collection for the list sample and the second contact for the non-overlapping area sample farms was done by telephone, saving time and resources. The collection of the survey data for both components at the same time was operationally feasible and reduced the burden on respondents.

In sections 3.2 to 7 of this chapter, the survey design of the Canadian National Farm Survey in the Province of Ontario is described.

3.1 NFS List Sample Design

The list frame for Ontario is made up of all 1986 census farms that either operated land in the province or had their headquarters within its boundaries. The list frame sample design is a single-stage stratified design with the farm as the sampling unit.

Stratification. For the NFS, all farms with unique operating structures - Hutterite colonies, multiholdings, community pastures and other significant farms (which stand out because of the large values of their crops, livestock or financial variables) - are identified and placed in a self-representing or "take-all" stratum so that they can be selected in the sample with certainty.

The very large farms are identified using an intuitive rule called the Sigma-Gap rule. A detailed description of how the rule is applied may be found in Trépanier and Théberge (1993).The Sigma-Gap rule was applied to the nine key variables to be estimated from this general purpose survey: 1986 Census values for total cropland, total expenses, total sales, beef cows, dairy cows, sows, total cattle, total pigs and total sheep. All farms for which any of the nine key variables can be identified by the Sigma-Gap rule are classified as very large farms and placed in the take-all stratum. Once the take-all stratum has been defined, the remaining farms on the list frame are further stratified and sampled.

The procedures FASTCLUS and CLUSTER of the SAS software (SAS Institute Inc., 1985) are used to determine the multivariate take-some strata by grouping farms that resemble each other or are close to each other in terms of Euclidean distance for the nine key variables. The FASTCLUS program groups the units to be stratified into 250 clusters. The mutually exclusive clusters identified using the FASTCLUS procedure are then combined to form the strata using the CLUSTER algorithm. In Ontario, with a population size of 72 599 farms, there are 81 strata, including the complete enumeration strata.

Sample allocation, selection and rotation. A multivariate allocation method with the same key variables as were used for the stratification is used to allocate the province's sample size to the strata. Once the initial stratum sample sizes have been calculated, they are adjusted to multiples of 4 so that four independent replicates may be chosen from each stratum that is not completely enumerated. The replicates are used to facilitate the calculation of the variance of the estimates and for rotating the sample.The list sample size in Ontario is 8 401 farms, of which 237 are in the complete enumeration strata and 8 164 are allocated to the remaining strata. Once the allocation of the sampling units to the strata has been completed, the units are sorted within each of the sampled strata by their subprovincial area and total operating expenses. Four independent circular systematic samples are chosen from this sorted list to form the four replicates.

The list sample design is fixed for five years. In order to reduce response burden, a 25% per year rotation rate is applied so that 75% of the sample would be common between years to enable good year-to-year change estimates. Self-representing units in the complete enumeration stratum are not rotated because of their importance to the estimates.

3.2 NFS Area Sample Design

The Population Census maps which show the location of census enumeration areas (EA) are the base for the area frame. Agricultural data are linked to the EAs because the Census of Agriculture is conducted together with the Population Census. The size of an EA will vary depending on the population density. Since farms can straddle EAs, all agricultural variables relating to a farm are ascribed to the EA which contains the headquarters of the farm. Thus, all EAs which have agricultural activity (i.e. contain the headquarters of at least one farm) within their boundaries are known and census data for key agricultural variables for the farms ascribed to them are available. These agricultural EAs define the area frame.

The sample has a two-stage stratified design with the enumeration area (EA) as the primary sampling unit (PSU) and a portion of the EA (segment) as the secondary sampling unit (SSU). Only EAs which have agricultural activity (i.e. farm headquarters located within the EA) are included in the area frame. To reduce costs, some remote regions of the country are excluded from the area frame even though they may contain a few agricultural EAs. Once all exclusions have been identified on the maps, the mapping team begins to outline the EAs which are in scope for the first stage of sampling. Ontario has 2 687 agricultural EAs in its area frame.

Stratification. Since estimates from the survey are required at the provincial and subprovincial area levels, the area frame for the province is first stratified by subprovincial area (SPA). The EAs within each SPA are grouped into homogeneous clusters based on 1986 Census data summarized at the EA level for the following key variables: total cattle, total pigs, total sheep and total farm area. These variables were chosen since they provide a composite measure of the agricultural activity of the EA based on both crops and livestock information. There are 49 strata in the area frame of Ontario.

First stage sample allocation and selection. After the stratification of the EAs, the first stage sample is allocated to each province and then to the subprovincial area within the province. As in the list sample design, the sample allocation at both steps is a compromise between the optimum allocation for each of the key variables. The SPA sample sizes are then adjusted to multiples of 8, in order to facilitate the creation of replicates. Strata with small sample sizes are adjusted to multiples of 2.

Before sample selection, the EAs within each stratum are sorted by their crop district and total agricultural sales. Two independent circular systematic samples of the

EAs are chosen within each stratum in order to form two replicates. Systematic sampling effectively implies stratification by crop district and total agricultural sales. Replication is necessary for variance estimation because often only one second stage unit (segment) is selected per PSU (EA). There are 195 EAs selected in the first stage area sample in Ontario.

Creation of segments within the selected EAs. Once the first stage EAs are selected, their boundaries are traced manually on topographical maps by the mapping clerks. These selected EAs are then divided into parcels of land, called segments, for the second stage selection. The guidelines require the segments to be roughly equal in size, with an area of approximately 6 to 10 km^2 of land and to be formed by following natural boundaries wherever possible. Based on the guidelines, the EAs are usually divided into 10 to 30 segments of roughly equal size.

Second stage sample allocation and selection. A simple random sample of segments is selected within each sampled EA using the sampling ratio of approximately 1 in 10. All tasks, from the identification of the primary sampling units PSUs (EAs) to the selection of the secondary sampling units SSUs (segments), are done manually by the mapping clerks who are given detailed instructions and training on how to select a simple random sample using random number tables. In some instances where the number of segments in a selected EA is greater than 20, two or more segments are selected within it keeping to the 1 in 10 sampling ratio. However, many EAs have less than ten segments and one segment is selected from each of them. In Ontario, 259 segments (SSUs) are selected from the first-stage sample of 195 EAs (PSUs).

4. NFS DATA COLLECTION

Statistics Canada has eight Regional Offices located in different parts of the country. The Regional Offices are responsible for the fieldwork of both the list and area components of the NFS. Personnel in the Regional Offices include the interviewers involved in collecting the data and their supervisors. None of the interviewers are permanent staff and, in general, they are women working part-time. Data for the farms in the list sample are collected using telephone interviews. These interviews are conducted in late May for information on crops and late June for information on livestock. During the May interview a "screening" question is included on the crops questionnaire to determine whether another contact is required in June to collect livestock data. Telephone interviewing is possible since the majority of households in Canada (98%, even in rural areas of Canada) have telephones. The telephone number is recorded on the census questionnaire and is therefore available on the frame.

All farms linked to the segments in the area sample are contacted through personal interviews in April or May to identify all land within the segment boundaries and to obtain their telephone numbers. Respondents operating land inside the boundaries of the selected segments may be contacted a second time in July to collect crops and livestock information. This second contact is done by telephone as part of the list sample data collection since the information required relates to the entire farm, and not just to the land inside the segment.

4.1 Supervisor Selection and Training

Supervisors are usually appointed from among past interviewers for the survey and are trained by the Regional Office staff and personnel from Head Office. The same home-study and classroom training exercises are completed by the supervisor and the interviewers. In addition, supervisors have extensive training in editing and follow-up procedures and on specialized situations which may occur in the field.

4.2 Interviewer Selection and Training

The selection and training of the interviewers for the area sample requires additional time and preparation compared with the list sample because of the personal interviews and the numerous enumeration procedures. Each year, an attempt is made to recruit interviewers who have previously participated in the NFS since these persons would require less training than new interviewers.

NFS interviewers are paid for their training time and are trained through the use of both home study packages and classroom training. Following the home study, the interviewers attend a half-day classroom training session given by their supervisor.

4.3 Data Collection for the List Sample

Once contact is made using the telephone number on the label of the questionnaire, the interviewer tries to speak with the respondent identified on the address label. If the operator is not available at the time of the interview, a proxy is allowed. The proxy is often the spouse or partner of the operator who tends to be well informed about the farm's operation.

After contact is made with the appropriate respondent, the information on the address label is verified to check for required changes. Following that, the interviewer obtains the survey data required on the questionnaire. This procedure continues until all respondents in the interviewer's assignment have been interviewed. The supervisor edits the questionnaires and if any of the edits identify errors or inconsistencies in the data, the supervisor contacts the respondent to correct the information. Remarks about refusals or failure to contact are noted on the questionnaire. The questionnaires are sent to the Head Office in batches for data capture.

4.4 Data Collection for the Area Sample

Data for the area sample are collected through personal interviews. The interviewers are sent into the field equipped with an "assignment package" that contains the Segment Control Register, master maps, assignment maps, segment maps and plotting maps. The Head Office mapping team is responsible for the preparation of this package.

Master maps. After the sample of segments has been selected, the segment boundaries are transferred from the EA maps that were used for sampling to another set of topographical maps (master maps). Only the segment boundaries are highlighted on these maps. Most of these maps have a scale of 1:50 000.

Interviewer assignments. The master maps are used in the Regional Offices by the supervisors to prepare the assignment for each interviewer (usually five to six segments). After the assignments for each interviewer have been determined in the Regional Office, the segments which make up each assignment are circled and numbered on the master maps. These maps are then sent back to Head Office where the mapping clerks check that the assignments are practical and then prepare the assignment maps.

Assignment maps. The enumerators' assignment map is a topographical map on which the boundaries of all segments to be enumerated by the interviewer are outlined clearly, together with other information such as the province, EA and segment identification number. The scale of the assignment map is 1:50 000 or 1:250 000 depending on the availability of the maps.

Segment maps. These maps are created for each segment and have a larger scale and more detail than the assignment maps to help orient the interviewer in the field. These maps can be town plans, county maps or topographical maps with scales which can range from 1:10 000 to 1:20 000 or 1:50 000 in the case of the topographical maps.

Plotting maps and segment control registers. A plotting map is prepared for each selected segment. These could be recent aerial photographs, low-scale maps, municipal or county maps with the lots identified on them. The scales of the maps or aerial photographs used for this purpose are usually 1:10 000 or 1:20 000. This map is shown to the respondent to determine whether or not the respondent operates land within the boundaries of the segment. A clear plastic overlay is fixed over the original map and the land operated by each respondent within the boundaries of the segment (tract) is marked or plotted on it.

In addition to the plotting maps, the interviewer uses a Segment Control Register to collect information on each agricultural tract within the boundaries of the segment. The interviewer enters the name and address of every farmer who operates some land inside the segment and lists the tracts of land operated in the segment.

All of the land within each segment, whether agricultural or not, must be accounted for using the plotting map. This information is used in estimation to prorate the data relating the total farm down to the segment. Each tract of land which appears on the plotting map has an entry on the Segment Control Register. Data required to calculate the estimation weights and information related to overlap detection are also noted on the Segment Control Register.

Enumeration. When a respondent who operates land inside the segment is found, the tract is outlined on the plotting map overlay. To this tract of land a serial number on the plotting map is assigned and the same number is used for the entry relating to that tract on the Segment Control Register. The interviewer then proceeds with the interview using the questions on the Segment Control Register. The full name of the respondent is recorded on the Segment Control Register and the interviewer determines whether the agricultural holding will receive at least $250 annually from agricultural sales related to the whole farm (not just the tract inside the segment). If the respondent does not expect to generate that amount from agricultural sales, the tract of land is marked as “non-agricultural” on the overlay of the plotting map and the Segment Control Register, and the interview comes to an end. If the respondent is eligible for the survey, the area (in acres) of his land operated inside the segment and woodland area within that tract is recorded on the Segment Control Register. These values are usually estimates provided by the respondent.

Once the information required for the plotting map and Segment Control Register has been obtained from the respondent, the NFS area component survey questionnaire, also called the area screening questionnaire, is completed to obtain information for use in detecting overlap between farms in the area and list frames. Then, the interviewer proceeds to the next operation within the segment; continuing until all land inside the segment has been accounted for. After a segment has been completed, the interviewer edits all forms for completeness and transmits the segment package to the supervisor. The supervisor completes any editing and follow-up that is required and transmits the segment package to Head Office for further processing, as described in section 5.

Figure 2.1 includes an example of a plotting map (municipal map) completed by the enumerator.

4.5 Overlap Detection

For the NFS, information from the complete list sample and only from the non-overlapping units of the area sample is used to calculate estimates using the screening estimator mentioned in section 1. The identification of sample farms common to both frames is therefore crucial. This is called "overlap detection".

The detection between the frames can be accomplished in the field at the time of data collection, or in a more centralized fashion after initial enumeration has occurred. The move of overlap detection to the Head Office, which allowed better control and verification, and the use of enhanced area frame screening questionnaires instead of screening books, has improved the overlap detection process which was previously carried out by interviewers in the field. If the farm is found on the list frame, information from the area screening questionnaire on the total area of land operated and number of livestock on the farm are used to verify that the match is a true match. Once all area sample farms are unduplicated against the list frame, the Regional Offices are informed about all non-overlap farms and these are contacted a second time together with the list sample to collect data on crops and livestock. Because data collection at this stage is by telephone, the interview costs for the non-overlap units are reduced.

5. NFS DATA PROCESSING

Although the data collection is done in the Regional Offices, all other data processing is centralized and occurs at Statistics Canada's Head Office.

5.1 Data Capture and Verification

The Segment Control Register data items are captured using the "heads down" technique on a Honeywell mini-computer system. No edits are applied during the data capture, but a 100% recapture of the key data fields is carried out for quality control. If there are discrepancies, corrections are made to ensure that the captured Segment Control Register data correspond to the actual data. No other changes to the data are made at this time.

The data items required for weighting and estimation are either on the Segment Control Register or on the crops and/or livestock questionnaire which collects data at the second contact for non-overlapping area farms. The data items from both the list and area sample crops and livestock questionnaires are captured in a similar fashion to the Segment Control Register data. There is also 100% verification of key fields on the questionnaire, such as total farm area, other crop areas and livestock number totals.

5.2 Daily Editing

After the data capture and verification has been completed for the Segment Control Registers and the list and area crops and/or livestock questionnaires, software specialized for this survey and other agricultural surveys, called the Integrated Agriculture System, is used to store and process the data on the mainframe. The Integrated Agriculture System, which is embedded in the ADABAS database management system, is used for editing, validation, retrieval and manipulation of the data. At each stage of editing, imputation or any manual updating, a "state" is assigned to indicate which step in the process has been completed for each data record. Any discrepancies related to the critical variables are documented and followed up. Manual imputation may take place at this point. It is preferable that errors be corrected at this stage rather than at the estimation stage.

5.3 Editing and Imputation

Following the daily editing process, additional edits are applied to identify partially missing sections of the questionnaire and inconsistencies between data items. In order to ensure consistency, deterministic imputation is performed in cases where a plausible value can be found for a data item that is missing or fails an edit.

Donor imputation is used for partial imputation when a section of the questionnaire has been refused or is partially completed and cannot be deterministically imputed. This process involves copying data that need to be imputed from a "donor" record that has passed all edits and that matches the "recipient" record that needs imputation. It uses the nearest neighbour approach to find the donor record based on five matching data fields that have been determined for each section of the questionnaire. These matching data fields have been determined based on subject-matter experience and on the correlation of the data items with the fields being imputed.

Records are used as donors if they have positive values for the data fields which require imputation on the recipient record and if they are the same farm type as the recipient. If an entire section of a questionnaire is missing, then all data items are copied from the donor record to the recipient record. When only some items require donor imputation, the donor is identified according to the matching of data fields and the donor's values are prorated before being copied to the recipient record so that totals agree. A total questionnaire refusal or non-contact is not imputed but is accounted for by adjusting weights of other respondents in the same stratum.

6. NFS ESTIMATION METHODS

The NFS estimates are calculated using data from both the list sample and non-overlapping area sample, which are completely independent. Estimates and their variances can be calculated for each sample separately and simply added together to give the overall estimate. Estimates of level and change from year to year are produced. Programs written with the Statistical Analysis System (SAS) are used to calculate estimates for the survey.

6.1 List Sample Estimates and Variances

Annual estimates of totals and their variances for the list sample are calculated using the usual stratified estimator:

$$\hat{Y} = \sum_{h=1}^{L} \frac{N_h}{n_h} \sum_{i=1}^{n_h} y_{hi}$$

where

L = Number of strata.
N_h = Total number of farms in the h^{th} stratum.
n_h = Number of sampled farms in the h^{th} stratum.
y_{hi} = Value for the variable of interest for the i^{th} sampled farm in the h^{th} stratum.

The estimate of the variance of $\hat{Y}$ is given by

$$\hat{V}(\hat{Y}) = \sum_{h=1}^{L} \frac{N_h(N_h - n_h)}{n_h(n_h - 1)} \sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)^2$$

In addition to those estimates, estimates of the year-to-year change ratio are computed and the variance of the ratio is calculated. Farms must be common to both years in order to be used for the change ratio estimate. Common farms are identified using the unique farm identification number (FARMID) that is given to each farm on the frame.

Thus the change ratio estimate is

$$\hat{R} = \frac{\hat{Y}_{current}}{\hat{Y}_{previous}} = \frac{\hat{Y}_c}{\hat{Y}_p}$$

and its estimated variance is

$$\hat{V}(\hat{R}) = \frac{[(\hat{R}+1)(\hat{V}(\hat{Y}_c)+\hat{R}\hat{V}(\hat{Y}_p)) - \hat{R}\hat{V}(\hat{Y}_c+\hat{Y}_p)]}{\hat{Y}_p^2}$$

6.2 Area Sample Estimates and Variances

The area sample estimates are calculated based on data for farms in the sampled segments that do not overlap with the list frame.

The formula used in the NFS to prorate the farm data is based on the appropriate fraction of land that is inside the segment, and is referred to as the *land weight* which is equal to the non-woodland area of the farm inside the segment divided by the non-woodland area of the total farm. In order to calculate the contribution of a farm in a sampled segment to the estimates, the data items collected on the questionnaire for the farm are multiplied by the land weight. The sum of the contributions of all farms in a segment provides the segment total for each data item.

The stratum totals are arrived at by summing the weighted segment totals within replicates, and averaging the replicate totals within a stratum. These are summed over the strata to calculate the overall totals:

$$\hat{Y} = \sum_{h=1}^{L} \frac{1}{R} \sum_{r=1}^{R} \frac{N_h}{n_{hr}} \sum_{i=1}^{n_{hr}} \frac{M_{hi}}{m_{hri}} \sum_{s=1}^{m_{hri}} Y_{hris}$$

where

L = Number of strata.
R = Number of replicates.
N_h = Total number of EAs in the h^{th} stratum.
n_{hr} = Total number of sampled EAs in the r^{th} replicate and h^{th} stratum (1st stage).
M_{hi} = Total number of segments in the i^{th} sampled EA of the h^{th} stratum.
m_{hri} = Number of sampled segments in the i^{th} EA of the r^{th} replicate and h^{th} stratum (2nd stage).
Y_{hris} = Segment total for the s^{th} segment in the i^{th} EA of the r^{th} replicate and h^{th} stratum.

The estimated variance is calculated using an estimator described in Wolter (1985, p. 19-26) and corresponds to a sample design with independent replicates.

$$\hat{V}(\hat{Y}) = \sum_{h=1}^{L} \frac{S^2(\hat{Y}_{hr})}{R} \quad \text{where} \quad S^2(\hat{Y}_{hr}) = \frac{1}{R-1} \sum_{r=1}^{R} \left(\hat{Y}_{hr} - \bar{\hat{Y}}_h\right)^2$$

$$\hat{\bar{Y}}_h = \frac{1}{R}\sum_{r=1}^{R}\hat{Y}_{hr} \quad \text{and} \quad \hat{Y}_{hr} = \frac{N_h}{n_{hr}}\sum_{i=1}^{n_{hr}}\frac{M_{hi}}{m_{hri}}\sum_{s=1}^{m_{hri}}Y_{hris}$$

is the estimate for the r^{th} replicate in the h^{th} stratum.

Estimates of the year-to-year change ratio for the area frame are based on segments which are common from year to year, and variances are calculated for these estimates using the replicate approach, as in section 6.1.

6.3 Overall Estimates and Variances

The overall estimates and variances for the estimates of totals are computed by simply adding the list and area sample estimates and their variances. The coefficient of *variation of the estimate for each data item is also calculated and included in publications* of the estimates at the provincial and national level. An adjustment is made to the survey estimates to account for the exclusions so that the final estimates relate to the target population. Domain estimates are not adjusted. The overall year-to-year change ratio is computed in a similar fashion, and variances are available.

6.4 Weighted Estimators of Segment Totals

The weighted estimator was chosen for the NFS because estimates for all agricultural items can be calculated precisely, and because the cost of enumerating all farms within a segment is not much greater than enumerating only some of the farms. When a weighted estimator is used, only data related to land inside the segment are collected at the time of the screening interview and less costly data collection procedures can be used for future survey contacts. The problem of underreporting total area has been addressed to some extent in Canada by excluding woodland from the land weight calculation. It has been observed that respondents tend to ignore woodland when asked for total farm area.

6.5 Raising Factor Adjustment and Weighting Problems

As in most surveys, there are cases of respondent total refusals or non-contact. In order to account for these situations, an adjustment to the survey design weights is made during the estimation process. Complete imputation of these records is not used because it is felt that weight adjustment is more efficient in terms of variance and provides a better representation of the non-responding operations. However, these estimators are subject to bias due to imputation.

Calculating the adjustments to the weights is done separately for the list and area samples. For the list sample, the weight within each stratum is adjusted by using the actual number of respondents in the stratum in the denominator of the raising factor, N_h/n_h , instead of the number of expected respondents based on the sample design. Once the stratum weights have been adjusted, the estimation process continues as usual. The adjusted weights are used in the variance calculations to reflect the true variance in the estimates. For the area sample, the number of usable segments is determined before any re-weighting occurs. If a segment is not usable, the weights of other segments in the stratum are adjusted accordingly. Segments with a legitimate zero value are considered usable. Since most farms are on the list, the number of farms which contribute to the area component of the estimates is small and can have such an impact on the weight that the final weights may be higher than the expected design weights. A computer program has recently been written which calculates weights for the area sample before estimation begins so that potential problems can be identified quickly.

7. NFS DATA ANALYSIS AND PUBLICATION

Analysis of the survey estimates by "subject matter experts" is a very important step in the survey process. At Statistics Canada each agricultural commodity is examined by economists working in the Agriculture Division who have specialized training and expertise in that commodity.

The purpose of this analysis is to ensure that the survey estimates are internally consistent at all geographic levels and farm type, as well as with other sources of information.

7.1 Analysis Using Administrative Sources

An important tool for analysing livestock estimates is the use of a supply-disposition balance sheet. A similar supply-disposition balance sheet involving stocks, harvested yields, imports, exports, feed and waste of grains is used to analyse crop estimates. In addition, for crop data administrative information is available from the Canadian Grain Commission and the Canadian Wheat Board. Details on the past three years of published estimates as well as the level and year-to-year change estimates are all incorporated in these balance sheets.

The five-year Census of Agriculture is also used as a benchmark for most agricultural items or products.

7.2 Analysis Using Survey Sources

A report listing the top 25 weighted contributors to the estimate is used by the subject matter experts to scrutinize any outliers or influential observations which may be present in the data. Another report describing the impact of imputation on the data is checked to ensure that imputation has not drastically changed reported values. After any corrections are made to the raw data, the estimates are tabulated again based on the updated information.

Once subject matter experts are satisfied with the preliminary estimates, this information is forwarded to the provincial statisticians for analysis and verification with locally available information (e.g. crop conditions) for their respective provinces. Changes suggested by the provincial statisticians may be incorporated into the estimates after further discussion, but the final decision on the estimates to be published rests with Statistics Canada's Agriculture Division.

7.3 Revision and Publication Dates

Crops estimates are usually published within three weeks of the reference date on the questionnaire. Crops data are subject to revision for up to two years after first being published.

The livestock estimates are published within approximately six weeks of the reference date on the questionnaire. The revision period for livestock data is usually within 18 months after first publication.

7.4 Dissemination: Products and Policies

Statistics from the Agriculture Division are available in a variety of formats including preplanned publications, customized tabulations and microdata. For special requests, data may be provided by telephone or by facsimile. The NFS estimates contribute to the two publications mentioned below. The *Field Crop Reporting Series* publication is released eight times per year at strategic times during the Canadian crop year. Livestock information is available through the *Livestock Statistics* publication which presents estimates four times a year at the provincial and national level for cattle and calves, hogs, and sheep and lambs.

8. THE AREA FARM SURVEY (AFS)

In 1993, the NFS was redesigned and replaced by separate surveys: the Crops Survey and the Livestock Survey. These surveys are also multiple frame surveys with separate list samples supplemented by a common area sample: the Area Farm Survey (AFS). It was felt that more efficient sample designs could be achieved by separating the two surveys. More important, the new design allows the collection of crops and livestock data at the time best suited to each of these variables.

Under the 1993 design, the crops and livestock surveys share the same list frame based on the 1991 Census of Agriculture. The samples for the crops and livestock surveys are selected independently, and this allows stratification and allocation to be optimized separately for the crops and livestock surveys.

Data collection for these list sample surveys was done in the Regional Offices using a Computer Assisted Telephone Interview (CATI) system. While on the telephone collecting the data from the respondent, the interviewer captured the information on a microcomputer and at the same time the data were edited. All of the daily edits mentioned in section 5.2 were incorporated into the on-line editing features of the CATI system.

The AFS was last conducted in 1995. In 1996 and 1997, the non-overlapping farms found through the 1995 AFS were used to account for new births (to 1995) and the 1991 Census undercoverage.

Whatever limitations existed with the AFS could probably have been solved by an increase in the sample size while at the same time decreasing the segment size. However, the current costs were starting to be prohibitive and any increase was deemed to be out of the question. Furthermore, the increased availability of administrative files, combined with advances in record linkage, was seen as an opportunity to steer the Agricultural Survey Programme away from a dual frame approach using a list and area component, to a dual frame approach using lists only.

At the moment, the income tax files are being studied as a potential source to identify new entrants to Agriculture and to account for census undercoverage by comparing the lists of tax filers on two successive years. A new Farm Update Survey will then be used to contact a sample of potential births to establish the status of farms linked to tax filers on 1996 Census Day. As with the AFS, a detection of overlap with the census list of farms will be undertaken. A second contact with non-overlapping farms will be done to collect survey data as before. A pilot Farm Update Survey is planned for September 1997; depending on the results of the pilot survey, the implementation is to take place in the spring of 1998.

Description of the AFS (taken from Théberge, 1994)

The AFS frame covers nine of the ten provinces of Canada, with the province of Newfoundland being excluded because of the relatively low levels of agricultural activity. In the Prairie Provinces (Manitoba, Saskatchewan, Alberta) and in part of the Peace River District in the Province of British Columbia, regular cells of 3 miles x 1 mile were created. These correspond to the fairly regular legal land description in use. Elsewhere, cells of 3 km x 2 km were constructed using the Universal Transverse Mercator grid. The digitized boundaries of the cells were then overlapped with those of the EAs using geographical information systems. The area of the resulting polygons was also calculated. Cells that do not overlap EAs with farm headquarters were automatically discarded. The farm headquarters is defined as the residence of the farmer when it is located on the farm, or else as the main farmgate. Other cells were discarded after being checked on a topographical map if they were completely inside a national or provincial park. At the same time, cells were combined when a large part of them was water. This combining of cells was coded, captured and a computer program ensured that the size of each group was reasonable, that the cells were adjacent and that they satisfied other desirable conditions. Each remaining cell, or group of cells if combining was done, corresponds to one of the segments that make up the survey frame.

The same subprovincial regions as those for the crop surveys were used in stratifying the segments of each province. A composite measure of agricultural activity based on the total numbers of cattle, pigs and sheep, total farm area and number of farms was computed to form strata within each subprovincial region. Statistics for the segments on the frame were arrived at by allocating EA census totals to the overlapping segments in proportion to the area of the overlap. The number of strata is such that there are on average 25 sampled segments per stratum. Stratification in the Province of Prince Edward Island was different and made use of digital satellite imagery to obtain more precise land use statistics for the segments. Images were available to cover this small province.

A sample of 2 100 segments was allocated to provinces roughly in proportion to their number of segments on the frame while taking into account the sample sizes under the previous design and respecting a minimum size. Sample allocation to the strata was proportional to the square root of the stratum size. Each stratum's segments were then sorted randomly within the census subdivision (usually municipalities) before systematic sampling was performed. To implement sample rotation, a second non-overlapping sample was selected in a similar manner and each of the two samples was divided into four stratified subsamples. Replacement of one of the subsamples can provide a rotation rate of 25%.

The sampled segments are plotted on topographical maps (scale of 1: 250 000) which are sent to seven Regional Offices for the formation of interviewer assignments. A computer program can graph the assignments as well as provide various statistics to help ensure the assignment quality. The interviewers are provided with topographical maps (scale of 1: 50 000) to help in locating the segments in their assignment. The interviewers must account for all the land within a segment. This is done by delineating the land of the various operators on an aerial photograph or a large-scale map. All farm operators with land in a selected segment are asked identification information which will help in the matching operation.Total farm area and dichotomous variables on the agricultural products of the farm are also collected. The area of the farm within the segment is measured by the interviewer with the help of a grid. In much of the Prairie Provinces and in the Peace River District of British Columbia where the regularity of land plots permits it, the interviews are conducted by telephone, a more cost-effective method than personal interviews.

The approximately 13 000 farms found within sampled segments are unduplicated against a register of known farms which comprises those enumerated in the 1991 Census of Agriculture. Most farm surveys use a subset of the 1991 Census farms as their list sampling frame. Farms of the AFS sample which cannot be matched to the list frame of a given multiframe survey are surveyed as part of the data collection activities of that multiframe survey in order to account for all farms not on the list frame.

The fraction of the total farm area inside the segment is used to arrive at segment totals. The estimates are then computed using the stratified (weighted) estimator based on the segment totals. Another estimation method using directly measured values for the segment totals (closed estimator) can be produced only for total farm area. Although closed estimators are more precise for seeded areas, they require collecting additional data for all farms in the area sample when only those not on the list frame are of interest. The relative importance of the Canadian list frames compared with the United States' list frames can explain the differing strategies.

The list frame of the Livestock Survey consists of all farms enumerated in the 1991 Census with the exception of farms on Indian reserves and institutional farms. The list frame is quite extensive when compared with those used in the United States, for example. The area frame of the AFS is used to identify farms that are not on the list frame either because of census undercoverage or because they have started operating after the census was conducted.

The estimates for each of the crops and livestock surveys are produced by adding the AFS estimates to the corresponding list sample estimates for each survey.

The Area Farm Survey is also being used to calculate multiple frame estimates for other agricultural surveys such as the Fruit and Vegetable Survey and the Farm Financial Survey. The AFS is an independent survey. For this reason, and since total land and the total land inside the segment are now collected on the Segment Control Registers for all farms in the AFS sample, it is also possible to provide estimates of the number of farms and total farm area based on the area sample only.

Table 2.1
1991 population distribution and number of farms in Canada

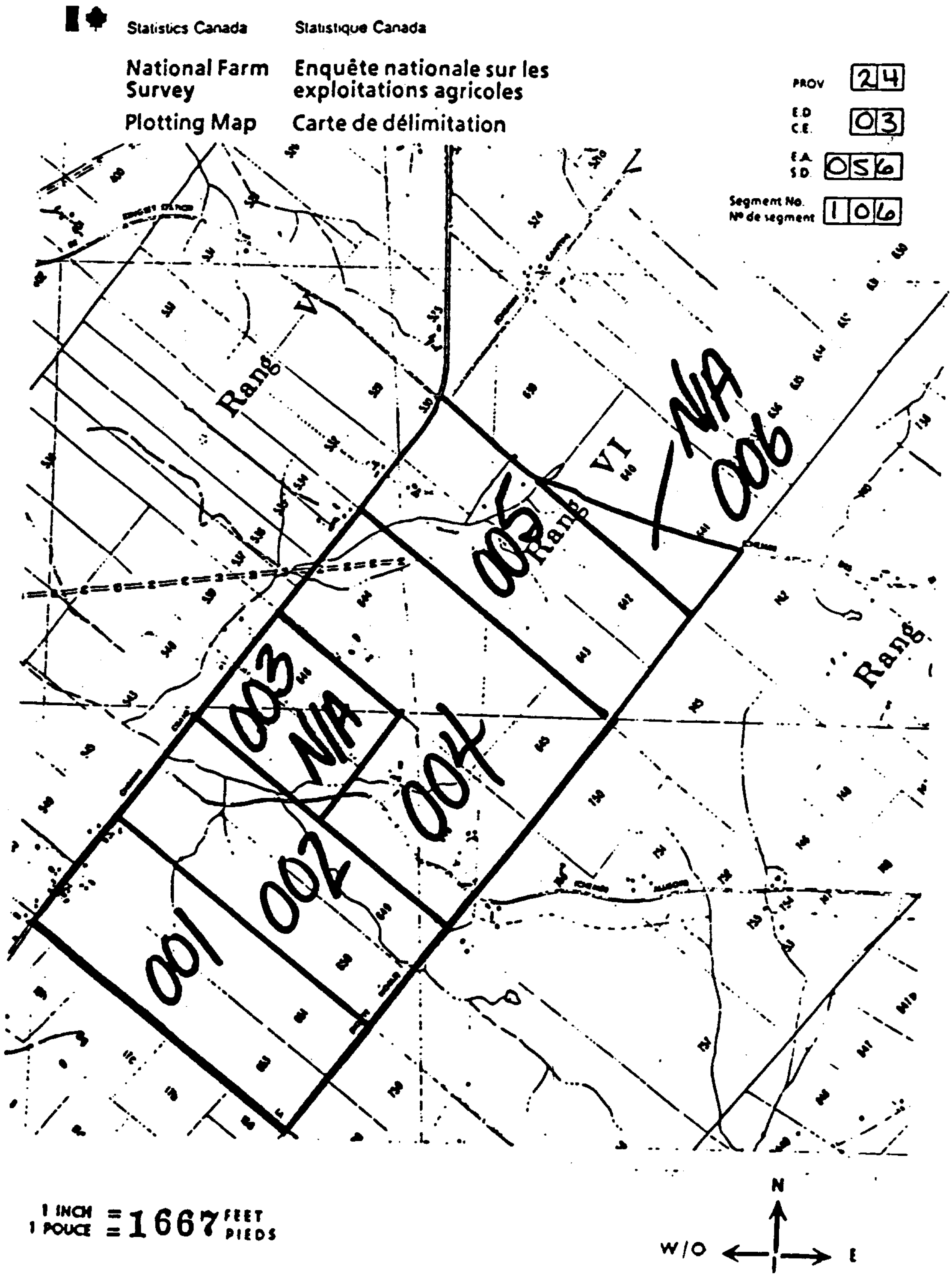
Province/ Territory	Population in 1991	Total number of farms	Farms reporting $2 500 (1) or more in gross farming receipts classified by farm type				
			Number of farms classified	Livestock farms	Grain and oilseed farms	Special crops farms	Mixed farms
Canada	**27 296 859**	**280 183**	**256 182**	**109 864**	**102 309**	**32 963**	**22 092**
Newfoundland	568 474	725	525	181	34	238	144
Prince Edward Island	129 765	2 361	2 144	1 216	593	224	222
Nova Scotia	899 942	3 980	3 300	1 860	95	1 195	300
New Brunswick	723 900	3 252	2 686	1 483	399	673	262
Quebec	6 895 963	38 076	35 600	22 755	4 420	7 446	1 958
Ontario	10 084 885	68 633	61 432	32 022	15 497	11 058	5 710
Manitoba	1 091 942	25 706	23 883	7 819	13 814	1 183	2 134
Saskatchewan	988 928	60 840	58 651	10 721	44 905	960	4 130
Alberta	2 545 553	57 245	53 443	25 587	21 214	3 872	5 540
British Columbia	3 282 061	19 225	14 518	6 220	1 338	6 114	1 692
Yukon Territory	27 797	113	---	---	---	---	---
Northwest Territories	57 649	27	---	---	---	---	---

NOTE (1): Only farms with receipts of $2 500 or more per annum are classified by farm type because classification by farm type is difficult for smaller farms.

--- *Not available.*

Figure 2.1

CANADA. National Farm Survey
Example of a plotting map (municipal map) completed by the enumerator



CHAPTER 3

HONDURAS

GENERAL PURPOSE AGRICULTURAL SURVEY PROGRAMME BASED ON MULTIPLE FRAME SAMPLING METHODS (1978-1997)

1. INTRODUCTION

The area frame in Honduras was constructed during the period 1978-1982 using uncorrected photomosaics for forming strata, PSUs and segments. Since then, the area sample has been used in many national surveys, particularly those dedicated to forecasting basic grain production (maize, beans, rice and sorghum) twice a year. The majority of agricultural land is used for the production of basic grains for internal consumption and for livestock feed (mainly poultry). Ninety percent of the farms produce or use these basic grains. These semi-annual surveys also cover other annual crops, vegetables, permanent crops such as coffee, cacao, and bananas, plus basic data on cattle, hogs and poultry.

The Area Sample Survey Programme was improved by the addition of the completely enumerated list of large holdings, thus becoming a Multiple Frame Survey Programme. The area sample has not been rotated and the frame has never been updated. There has been a lack of technicians, lack of budget and strong internal disagreement over the proper regionalization of the frame.

2. PRIMARY AGRICULTURAL CHARACTERISTICS OF HONDURAS

The agricultural sector supplies about 26% of the gross national product. The rural population constitutes 57% of the total population of 5.1 million. Agricultural and natural resource products make up around 80% of the country's exports. The most important export items are bananas, coffee, shrimps, wood and beef. With agricultural diversification, other crops such as citrus, vegetables and other tropical fruit are becoming important.

Of the 112 490 km^2 in the country, approximately 35 000 km^2 are dedicated to agriculture (not including forests). The majority of agricultural land is used for the production of basic grains (maize, beans, rice and sorghum) for internal consumption and for livestock feed (mainly poultry). Ninety percent of the farms produce or use these basic grains. According to a forest map made on satellite images in 1990, 66% of the country is forest land. The other 34% includes agricultural zones, pastures, rivers, lakes and other land. The composition of the forest on the forest land is 38% conifers, 32% broad-leaved trees and 30% deforested with slopes generally more than 30% in the deforested areas.

Most of the coffee is shade grown (cultivated under broad-leaved trees), making it almost impossible to identify on images or photography.

The livestock industry of the country can be characterized as *extensive*. It occupies most of the recently deforested areas. The desire for additional pasture is contributing to the rapid deforestation.

Small subsistence farms make up a high percentage of all farms. These small farms produce virtually all of the basic grains and vegetables. Gathering reliable data from the small farms is difficult because of a high rate of illiteracy, a lack of records and the use of varied local units of measure.

Administratively the country is divided into 18 departments which are, in turn, divided into 292 municipalities. The survey design was devised to provide estimates for seven agricultural regions covering the entire country. Unfortunately, the regionalization does not coincide, in many instances, with departmental and/or municipal boundaries. In fact, the regionalization utilized by the agency in charge of the area sampling frame, the Secretariat of Planning, Coordination and Budget (SECPLAN) is not the same as that employed by the agency in charge of the agricultural sector, Secretariat of Natural Resources. These differences are a major problem for the development and general acceptance of an agricultural statistical system.

3. SURVEY DESIGN

3.1 General Characteristics

As already mentioned, the general purpose Area Sample Survey Programme was improved by the addition of the completely enumerated list of large holdings, thus becoming a Multiple Frame Survey Programme.

The area sample consists of a replicated, systematic two-stage sample design - a systematic selection of CUs with probability proportional to size measures (PPS) and a simple random selection of segments within the sampled CUs - in which strata, CUs and segments have recognizable physical boundaries (cf. appropriate sections of Volume I of this publication).

The area frame is stratified according to intensity of land use. The list frames are stratified by type and size of farm.

3.2 Construction of the Area Sampling Frame

3.2.1 Stratification

Stratification, as mentioned above, was based on the percentage of land under cultivation in a zone or block of terrain. Land was classified as cultivated if it is used to grow annual crops such as basic grains, vegetables, cotton and/or permanent crops such as coffee, bananas, cacao. Land without agriculture or livestock such as national parks, military posts, swamps as well as large bodies of water were delineated and measured in order to account for total area of the country. The stratum of "projected water" was for areas that were to be covered with water in a future building of a dam. This made it possible to sample before the dam and easily remove the sampled strata when the dam was built. Cities were broken into two strata: one was the central portion of large cities; the other around the outskirts of large cities and in the small villages where a high density of population existed together with small plots of cropland.

Stratum I was divided into three strata with the same intensity of cultivation but including farms with different characteristics and, therefore, can be separated to make special estimates, if desired, and reduce overall variance. One subdivision was made to focus on coffee, another for irrigation and the last to include all other land that is 60% to 100% cultivated.

Where deforestation is rapid, as it is in Honduras, it was found to be better to extend the agricultural strata into what appeared at that time to be forest in areas where rapid deforestation was taking place. This procedure generally increases the length of time a frame can be used without being updated. The results from the frame might be less efficient at the beginning but would improve as deforestation continued and the land that was forest became cropland. (See Table 3.1 for stratum definitions)

3.2.2 Area Frame Construction Materials

Aerial photography. The land use stratification was constructed on aerial contact prints in a scale of 1:20 000. An enlargement from these prints to a scale of 1:5 000 was made for each sample segment to guide the fieldwork.

Topographic maps. Topographic maps in a scale of 1:50 000 formed the base of the area frame. These were used for the measurement of the counting units.

3.2.3 Delineation of Strata

Simple non-corrected photomosaics were constructed from the aerial contact prints for defining the strata, counting units and segments (for details on mosaic construction see Volume I). Satellite images were not available for area frame construction.

The next step after constructing the mosaics was to draw the national and regional boundaries on the mosaics. National boundaries were drawn on the mosaics exactly as they appeared on the official maps. Regional limits were delineated following physical boundaries that can be found on the ground. The official regional boundaries were often straight lines or imaginary limits that have no relationship to identifiable boundaries that can be found on the ground. The process was to find physical boundaries as near as possible to the official line, balancing land taken from the region with some added in at a later opportunity.

Once the national and regional boundaries were drawn on the mosaics, the stratification was carried out within regions on the mosaics using coloured grease pencils.

3.2.4 Counting Units

Following stratification, the strata were divided into counting units (CUs). A CU was a block of land with easily recognizable boundaries containing, on average, an area equivalent to six to ten segments. The size could vary from a minimum of two segments to a maximum of 15 segments. Segment size varied according to the stratum. CUs were formed so that each one represented the stratum to which it belongs. The formation of CUs provides a means of selecting a sample of area segments without subdividing the entire frame into segments (cf. Table 3.1 for CU and segment sizes).

All boundaries were transferred from the mosaics to the topographic maps. The CUs were numbered in a serpentine manner in each stratum within each region. Next, the CUs were measured with a planimeter and grid (cf. Volume I for details) and listed with their measured area and identifying number. A number of segments was then assigned to each CU by dividing the area of the CU by the target size of segment for the stratum. Segments were assigned integer numbers.

3.2.5 Selection of Replicated Samples

An accumulative column of assigned segments was prepared on the CU listing sheet for each stratum. The total sample allocated to the stratum was divided into the desired number of replicates. The size of one replicate was then divided into the total number of segments from the cumulative column to provide a sampling interval. From a random start, the sampling interval was applied systematically to select the first replicate. Another random start was used to select the next replicate, and so on. Selected CUs were subdivided into the assigned number of segments and one segment was selected at random. Sample segments were given an identification number that would locate it according to region, department, stratum and replicate plus a segment order number (cf. Table 3.2 for the total number of segments assigned to each stratum by region and the sample size).

3.3 List Frame Component

The list frames being used in Honduras consist of the names of large producers of livestock and basic grains (mainly rice) and other crops such as cantaloupes, watermelons and bananas. The lists are small and are completely enumerated during each survey. Farmers surveyed with the area sample are checked against the list sample. Those on the list are removed from the area tabulations.

4. ESTIMATION PROCEDURES

4.1 Estimators

Honduras has used both the open segment and closed segment estimators with the area frame surveys. The closed segment estimator has been used for forecasting and validating area and production of basic grains. Good results have been obtained for maize and beans. Results for rice have not been so good because this crop is concentrated in a few localities. The open segment estimator has been used for general purpose surveys where entire farm data are required. It also provides an estimate of number of farms and allows publication of data by size of farm. The open segment estimator is also used for livestock. The weighted segment estimator has not been used in Honduras.

4.2 Estimated Totals and Variances

Estimated totals and their variances are calculated by the usual procedures and formulas employed with stratified, systematic, replicated sampling. The formulae used can be found in Chapter 11.

5. DATA COLLECTION

5.1 Selection and Training of Field Personnel

It was found that the administration of a short test was a good means for choosing potential field staff. The test covered ability to understand and interpret written instructions, map reading and simple mathematics. Those who passed the test were then given a short course covering basic cartography, questionnaire content, interviewing techniques, questionnaire review and field supervision. The candidates were then evaluated by those who presented the course. Those who were lazy and/or showed little ability to learn were rejected. Those selected to be enumerators moved into formal training that included home study, group training and field practice. A comprehensive Enumerator's Manual was used for the home study. In the beginning, the training was conducted on a regional basis. However, the Honduras experience indicates that the quality of training is better when carried out at the central office.

5.2 Survey Materials for Field Supervisors

- Topographic maps showing the segments under their supervision.
- Extra questionnaires and segment control sheets.
- Blue and red grease pencils for drawing tracts and fields on the photos; pencils, pencil sharpeners, erasers and a clipboard.
- Plastic grids in *manzanas* (0.7 ha) and hectares for measuring areas on the photos.

5.3 Survey Materials for Enumerators

- Questionnaires and segment control sheets.
- Photo enlargements in a scale of 1:5 000 for each segment to be enumerated.
- Red and blue grease pencils for delimiting tracts and fields.
- Plastic grids for measuring areas on the photos.
- Pencils, pencil sharpener, erasers, clipboard, and plastic bags for the enlargements.

5.4 Organization of Fieldwork

Each survey is carried out by the following field personnel: a chief of field operations; seven regional leaders; 30 supervisors; 90 enumerators; and 32 chauffeurs with vehicles. With this level of personnel, it takes approximately two months to complete a national survey covering 1 190 sample segments. The delay is acceptable because of the difference in crop development between regions. One general purpose national survey is done each year.

The collection of correct data by the enumerators and supervisors is of the utmost importance because once the questionnaire arrives in the central office there is no possibility of contacting the farmer a second time to make corrections. Errors discovered in a review of the questionnaire by the enumerator and/or the supervisor in the immediate locality of the segment are the only errors that can be corrected with actual respondent data. Corrections made in the central office are based on averages or imputations from neighbouring farms. Telephones are few and far between in the field and the mail service is not reliable.

5.5 Questionnaires and Survey Forms

In all surveys, the determination of the number and type of variables to the investigated requires a balance between the need for the information and the cost of gathering data while maintaining an effective communication with principal data users. The *quality* of the data gathered is normally inversely proportional to the *quantity* gathered (cf. Volume I). The development of a questionnaire should start with a view of the end product and the preparation of the output tables that are expected by the principal data users. The questionnaire design should then be used in a pretest in all areas where it is expected to be used.

A *Segment Control Register* and the *Questionnaire* are the only forms used in Honduras.

5.6 Field Procedure for Data Collection

Once the supervisors have laid out their route and calendar of activities, the first action on arriving at a segment is to identify the boundaries. The enumerator must go around the segment by vehicle or on foot to be sure that all boundaries are recognized. When the segment is clearly identified, the enumerator proceeds to the nearest occupied dwelling within the segment in search of a farm operator. In the rare case where there is no dwelling in the segment, a search is made for a holder or qualified informant at work inside the segment or in some bordering area. When a farm operator is found with land inside the segment, an interview begins.

The land operated by the farmer within the segment (tract) is drawn on the photo with a blue grease pencil. Each of the person's fields are outlined with a red grease pencil. The segment control register is completed to determine if the operator is to be assigned a questionnaire. If so, the interview continues. The closed segment portion of the questionnaire is completed first and, if the operator qualifies by having his dwelling in the segment, the open segment portion is completed. It should now be possible to gain information on another operator with land in the segment and proceed to the next interview.

The review of completed questionnaires in the locality of the segment is possibly the most important work of the supervisor; attempting to catch and correct all errors and inconsistencies before the questionnaires are forwarded to the central office.

The operators on the list frame are enumerated by a team of experienced enumerators that is totally independent of the area frame. Contact with the large operators is considered to be sensitive and important.

6. DATA PROCESSING

6.1 Review, Editing and Coding

The review and editing of questionnaires must begin as soon as possible after the survey is under way, preferably by the supervisor in the field where data collection is being done. In this way, misunderstandings can be corrected and poor enumerators can be replaced. Edited and corrected questionnaires are sent to the central office for further checking and preparation for data entry.

6.2 Data Entry and Summarization

Each edited and coded questionnaire is entered into the computer and then verified (re-entered). An experienced data editor is on hand during data entry to correct any errors that appear at that stage.

The Central Statistical Office provides 15 terminals connected to a Wang-VS 100 computer for data entry. This same computer is used for summarization of the survey data, calculation of variances and file management. The necessary software was prepared by Honduran technicians using COBOL programming language.

7. CONCLUSIONS AND THOUGHTS ON FUTURE NEEDS FOR THE NATIONAL AGRICULTURAL SURVEY PROGRAMME

- In spite of the fact that multiple frame surveys have been carried out in Honduras since 1984, it has not been possible to establish a continuous system of agricultural surveys that is institutionally stable and consistent.
- Effective communication should be organized with principal data users. Under the existing situation, it has been difficult to set data priorities as well as evaluate what has been undertaken.
- The last National Agricultural Census of 1993 should be used to improve the overall survey design. The census can provide a new list frame for the Multiple Frame Survey Programme.
- Having a replicated sample has not been properly exploited. There has been no rotation of segments since the surveys began.
- Implement adequate quality control over data collection. Politics has adversely affected selection of survey personnel.

- The publication of results has not been timely and therefore not particularly useful for making decisions. A calendar for the publication of survey results must be established and followed.
- A cost-benefit study should be done to decide if it is worth the effort to continue obtaining regional estimates or if the sample should be modified to provide estimates only at the national level. If regional estimates are to be continued, a single regionalization that conforms to real differences in agricultural characteristics must be agreed on. The estimates to date from the area frame sample have been produced with a regionalization that is different from that of the Secretariat of Natural Resources.
- Improve the capability to analyse and evaluate survey results. A detailed analysis of survey results prior to publication merits the undivided attention of experienced agricultural statisticians and economists with sufficient time to complete the work to their satisfaction. These experts would assemble all known and available check data, census data and results from previous surveys and surveys conducted by other agencies for comparisons and verification. These experts might recommend : i) that the heavy use of imputations in certain areas be explained; ii) that certain estimates with a high variance not be published or be published with a clear description of their limitations; iii) that survey procedures and/or questionnaires be altered for future surveys; iv) that changes be made in editing, coding and software if any one of these is found to be contributing to discrepancies; and v) that a preliminary publication be made without the questionable items if it seems that considerable time will be involved in making corrections.

Table 3.1

Honduras: stratum descriptions with counting unit and segment sizes

Stratum	Stratum Description	Counting Unit Size (km^2)	Segment Size (km^2)
I_1	60-100% cultivated	4-6	0.5
I_2	60-100% cultivated with at least 30% coffee	4-6	0.5
I_3	60-100% cultivated with irrigation	4-6	0.5
II	30-59% cultivated	8-12	1
III	0-29% cultivated	20-30	2
IV	Potential for agriculture but not used	32-48	4
V	Non-agricultural land	20-30	2
VI_1	Urban centre - no agriculture	6-10 blocks	1 block
VI_2	Agro-urban (high population density, some agric.)	6-10 blocks	1 block
VII	Large bodies of water	Not sampled	
VIII	Projected water - divided among Strata I, II and III as appropriate		

Table 3.2
Honduras: number of segments in the population and in the sample by region and stratum

Region	Stratum															
	I_1		I_2		I_3		II		III		VI_1		VI_2		Total	
	N	n	N	n	N	n	N	n	N	n	N	n	N	n	N	n
1	800	30	148	8	896	28	1 020	25	6 560	20			1 876	12	11 30	123
2	228	27			512	32	330	24	19 836	66			1 817	12	22 72	161
3	694	32			2 800	66	946	50	25 452	88			6 393	20	36 28	256
4	1 536	54			866	25			11 184	55			1 794	8	15 38	142
5	1 340	45			168	20	379	24	57 204	79			1 731	10	60 82	178
6	1 306	42			188	16	233	16	24 696	88	183	6	6 264	18	32 87	186
7	340	20			178	12	846	28	14 580	72			2 360	12	18 30	144
Total	**6 244**	**250**	**148**	**8**	**5 608**	**199**	**3 754**	**167**	**159 512**	**468**	**183**	**6**	**22 235**	**92**	**197 684**	**1 190**

NOTE: N = Total population of segments in the stratum.
n = Number of sampled segments.

CHAPTER 4

UNITED STATES

GENERAL PURPOSE AGRICULTURAL SURVEY PROGRAMME BASED ON MULTIPLE FRAME SAMPLING METHODS (1945-1997)

COMPUTER-ASSISTED AREA FRAME CONSTRUCTION AND SAMPLE SELECTION (1993-1997)

1. THE AGRICULTURAL SURVEY PROGRAMME BASED ON AREA AND MULTIPLE FRAME SAMPLING METHODS (1954-1997)

The National Agricultural Statistics Service (NASS) of the United States Department of Agriculture (USDA) through its network of State Statistical Offices has the responsibility of providing the official agricultural statistics of the United States.

Statistics developed by NASS/USDA are obtained mainly from a variety of sample surveys and are essential for the orderly development of production and marketing decisions by farmers, ranchers and other agricultural businesses. These agricultural data series are also used for monitoring the ever changing agricultural sector and for carrying out agricultural policy relating to farm programme legislation, commodity programmes, agricultural research, agricultural chemical usage, rural development, environment and related activities.

Area and multiple frame probability survey methods provide the statistical foundation for conducting the agricultural probability surveys based on complete coverage of the United States agricultural sector.

Since the mid-1950s, NASS research activities on development of area and multiple sampling frame agricultural survey methods constitute a very major contribution in the world to large-scale agricultural survey methods and practices.

NASS has accomplished significant progress in its large survey programme that provides current estimates for hundreds of agricultural commodities. The following points indicate some of the research and development efforts:

- The introduction of probability survey sampling methods and practices, and quality control procedures for data collection, processing, analysis and data dissemination, have had a decisive impact on the reliability of US agricultural statistics. Until the 1960s, almost all US agricultural statistics published by USDA were based on non-probability surveys. In the 1960s area and list probability sample methods were introduced and began replacing many non-probability surveys to support national statistics for the agricultural sector of the United States. *After years of research and development work, a major part of US agricultural statistics are currently based on probability sampling methods.*

- Computer technology has been applied to a larger extent than in any other country for statistical modelling and analysis, data processing and many aspects of survey research and development. NASS started using computers in 1962.
- The pioneer application and integration of digital cartography, aerial photography and satellite imagery (for purposes of area frame construction and area sample selection) in the probability survey models have resulted in successful practical programmes.

The *Multiple Frame Agricultural Survey* is a general purpose operational programme that was initiated in 1986 in 27 states and later extended nationwide. The Multiple Frame Agricultural Survey is conducted in June (called the base survey or June Agricultural Survey [JAS]), July, August, September, January and March for estimating crops (area, production and planting intentions), livestock, land in farms, number of farms and grain and rice stocks.

The field data collection for the area frame component of the follow-on surveys (after the June base survey) are based on subsamples of the JAS area sample frame. Beginning in 1997 there will be an independent area sample for the fall survey (December). The area sample component of the June Agricultural Survey constitutes a very important tool for many other agricultural surveys in the United States; for instance, for the Monthly Objective Yield Surveys, the Agricultural Labor Survey and the Environmental Chemical Use Surveys.

In this section, the June Agricultural Survey (JAS), the largest annual survey conducted by NASS, is considered. The JAS area frame covers all the United States except Alaska and Hawaii (48 states). The JAS sample of the 48 conterminous states is currently formed by approximately 11 000 segments.

The area frame consists of the combination of the area frames constructed independently for each state which are updated every 10 to 20 years.

The surveys collect data on: planted crop areas and areas to be planted, land use areas, crop yield and production, irrigated cropland, grains and oilseeds in storage, number and land in farms, production of agricultural commodities, livestock inventories and values, pigs farrowed, calves born, land values and rents, cost of production, farm expenditures, farm labour, and other economic and social characteristics of the agricultural operations.

The JAS data collection is undertaken during two weeks and involves 1 600 enumerators. The enumerators conduct face-to-face interviews with all farmers who operate land in the selected segments. With the respondent's assistance, the enumerator identifies and delineates field boundaries in the aerial photography. Enlarged aerial photography of each sample segment (in a 1:8 000 scale) is used by the enumerators to facilitate field data collection. A farm is defined, since 1975, to be all land under one operating arrangement with gross farm sales of at least US$1 000 a year. From 1955 to 1974, a farm was defined as being all land under one operating arrangement with 10 acres (4 ha) or more and sales of US$250 a year or more without regard to acreage.

Two direct expansion area sample estimators are currently used: the *closed segment estimator,* for the main crop areas, and the *weighted segment estimator*, for all other variables.

In addition, ratio estimators are also used. The use of *open segment estimators* was discontinued, and replaced by the weighted segment estimator (cf. e.g. Chapter 11 or Volume I, Chapter 4).

Large-scale list frames are regularly updated and used in combination with the area frame. Of the approximately 2 million farms in the United States, the list frames cover around 1.1 million farms, and the list frame samples include approximately 70 000 farms, of which 10 000 are extremely large operators. On the other hand, the 11 000 area sample segments include approximately 37 000 farms.

Multiple frames are used, in particular, because for many agricultural commodities a high percentage of the total production is concentrated in a small number of large farms that cannot be properly represented by a relatively small area sample. Therefore multiple frame estimators are necessary. For example, there are 157 000 hog farms, and around 670 farms with an inventory of more than 1 000 hogs account for 67% of the total hogs.

Sampling errors for major agricultural items average about 4% to 8% at the State level, about 2% to 3% at the regional level and 1% to 2% at the national level.

Brief history of the application of area and multiple frame sampling methods

The first two large-scale (statewide) tests of probability area sampling for agriculture in the United States involving the measurement of crop areas in small segments were designed and conducted in 1938 and 1939 by the Iowa State University in cooperation with the Bureau of Agricultural Economics/USDA. At the same time, analogous survey methods were applied in India, apparently independently. By 1945, area frames (called Master Samples) were developed for all states of the United States, in cooperation with the Bureau of the Census (BUCEN).

In 1954, NASS began investigating the use of area sampling by conducting experimental surveys in ten states (703 segments) to evaluate estimates of crop areas. These area sampling frame surveys were geographically expanded over the years and made operational for the first time in the 48 conterminous states of the United States (all the United States area except the States of Alaska and Hawaii) in 1965, providing estimates at the state, regional and national levels. The survey was gradually refined and expanded in content so that a larger number of variables were studied. In 1967, the survey programme provided area frame estimates for crops, livestock, grain stocks, farming systems, as well as economic and social characteristics of the farms. The survey programme became a general purpose survey programme. The statistical foundation of the programme is best described in Houseman (1975) and Huddleston (1976).

The theory of multiple frame survey sampling and estimation was developed in the early 1960s, when NASS supported research at Iowa State University under the leadership of H. O. Hartley. Multiple frame methods were first evaluated in 1965. An important improvement in the Agricultural Survey Program was the introduction of multiple frame survey methods, beginning in 1969, to estimate livestock. By late 1979, multiple frame survey methods were regularly used to estimate cattle and livestock inventories, farm production expenditures, number of farms, farm labour and area for special crops. It should be mentioned that during the 1970s research and development efforts led to the construction and maintenance of large-scale general purpose list frames for all states for use in multiple frame and non-probability mail surveys.

Area frame replicated sampling designs have been used since 1974. Point sampling procedures were utilized for many years in the range stratum and the non-agricultural stratum of several states.

Concerning the development of area frames, from 1954 to 1992 aerial photography (in combination with paper maps) was the basic cartographic material used for area frame construction. Since 1978, satellite images have been the primary cartographic material used to identify land use strata. However, until the advent of Computer-Aided Stratification and Sampling (CASS) in 1993, strata boundaries continued to be determined on aerial photography and transferred to maps for digitizing and sample selection. Since 1993, digital satellite images in combination with digital maps have been used as the primary cartographic material used for area frame construction and sample selection.

The Landsat space programme provided the essential tools for the development of the use of satellite data in agriculture. NASS was a principle research partner with the National Aeronautics and Space Administration (NASA) and the National Oceanic and Atmospheric Administration (NOAA) in remote sensing applications to agricultural statistics. The partnership began in 1972, in concert with the launching of Landsat I in July of that year. There have been five Landsat satellites in the series to date.

An important improvement for area frame construction was the utilization of satellite images for the construction of strata and primary sampling units (PSUs). Satellite images provided more recent data (within two weeks of the satellite overpass) than the aerial photography used (may be five to ten years old). Owing to the fact that an area frame is utilized for about 15 years, having the most recent imagery at the time of stratification and PSU construction provided an advantage.

NASS first used satellite images printed on paper in 1978 (Hanuschak and Morrisey, 1977). The images were from the Landsat Multispectral Scanner (MSS) - aboard the first Landsat satellites - with a resolution of 60 m on the earth's surface. The images were printed only on paper in the 1:250 000 scale. Since 1979, such satellite images were used for area frame construction and partially substituted aerial photography. The satellite images gave a more up-to-date view of the land to be stratified, but lacked the detail to identify accurate frame boundaries. In 1982, Landsat IV was launched. This satellite contained the Thematic Mapper (TM) sensor with a resolution of 30 m on the earth's surface. The TM images were still printed on paper at the 1:250 000 scale and a few years later NASS substituted the MSS images for area frame construction. In addition, the series of French SPOT satellites (the first launched in 1985) provided even higher resolution images and digital data, beginning in 1986. The SPOT satellites provide black and white data with a resolution of 10 m on the earth's surface. These data are especially useful in urban and other critical areas where more detail is needed. The SPOT satellites are pointable, which increases the probability of obtaining cloud-free coverage. However, an agreement must be made to point the satellite well ahead of the time of overpass. An important disadvantage is their cost. SPOT data cost to cover the same area is approximately nine times that of TM data.

When obtaining satellite images in areas where clouds are a problem, a compromise must sometimes be made between the most recent date and earlier dates with less cloud cover. Moreover, because of their resolution, TM satellite imagery does not substitute completely for aerial photography (2m resolution), although it has done so to a very large extent. In particular, physical boundaries and urban areas cannot be seen as well in TM data, so the analyst must resort to using such characteristics as density of roads. For area frame construction, aerial photography or higher resolution satellite data is used to supplement TM data when higher resolution is needed, as in the case of urban areas. Also, sample selection in urban areas is being done on quadrangle maps (1:24 000 scale). New satellite technology in the United States, Canada, France, Russia, Japan and India is currently being studied to evaluate its application for improving agricultural statistics.

In 1987, NASS, in cooperation with the NASA/Ames Research Center, started to develop the Computer-Aided Stratification and Sampling CASS software, an image software system that employs digital processing of satellite and geographic attribute data to facilitate area frame construction and sample selection. The CASS software was designed to automate the manual methods used for area frame construction and sample selection based on cartographic paper products, as described in Volume I.

In 1993, the first area sampling frame and sample selection for a state (the State of Oklahoma) using exclusively CASS was undertaken. This state sample was used in the 1995 June Agricultural Survey.

Since 1993, CASS has been implemented into the NASS operational survey programme, and is now the only system being used in the United States for area frame construction. At this time, sampling frames have been prepared with CASS for the following nine states: Oklahoma (constructed in 1993), California (1994), New York (1995), South Carolina (1995), Ohio (1996), Kansas (1996), Maryland (1997), New Jersey (1997), Delaware (1997); and the Texas area frame is currently being constructed.

CASS constitutes the most important step in the series of improvements for area frame construction and sample selection during the last ten years.

2. COMPUTER-ASSISTED AREA FRAME CONSTRUCTION AND SAMPLE SELECTION (1993-1997)

This section briefly describes the procedures for developing area frames using digital inputs adopted by the National Agricultural Statistics Service (NASS) of the US Department of Agriculture (USDA) since 1993.

These procedures are highly automated and the material requirements of hardware, software and digital satellite data are difficult to meet in most other countries. However, these methods are important because they are employed at present in the largest and oldest area frame project in the world and because they represent the trend of the evolution of area frame construction and sample selection from manual to computerized procedures.

This section is derived mainly from three papers: Cotter and Tomczak (1994); *CASS*, a publication of NASS; and Dodson (1996) with updates from technicians currently working with CASS.

The utilization of land is always changing. As a result, the frame "ages" and the survey statistics begin to decrease in reliability. With the current availability of resources, only two new state frames (on average) can be developed in a year. Therefore, frame development is prioritized according to need. States are selected to receive new area frames based on a statistical analysis of survey results over time and on the condition of the old frame material. The manual procedure used to develop area frames was very labour intensive. The development of an area frame on paper-based cartographic materials for a typical or average US state may require 11 000 staff hours and cost over $150 000. Computer-Aided Stratification and Sampling (CASS) has cut the time requirement by well over one-half.

2.1 The CASS Workstation

A UNIX-based Hewlett-Packard (HP) workstation is being used to handle data processing and storage requirements. The graphics screen in the HP workstation possesses the minimum capabilities for area frame development, that is, three image planes, four overlay planes and a 1024 by 1280 display coordinate system. The image planes display three bands of satellite data using 24-bit colour. The graphics overlay planes are used for various purposes, such as displaying digital road, water and county boundary data and delineating and displaying PSUs for a county and surrounding area. A terminal screen is used to display menus (commands, etc.).

The keyboard is used to enter commands or filenames, and the mouse is used to create or edit polygons (PSUs and segments) and to vary the intensity of colour in the image planes. The function keys are also used to turn the overlay planes on and off, to change the colour of the overlay planes and to change the scale of the image on the graphics screen.

2.2 The CASS System

The CASS software is an image analysis system which incorporates digital imagery, digital line graph data and digitizing functions. As developed originally, the CASS software worked only with a specific type of hardware. The software is now being modified, making use of commercial software, so that it will operate with other hardware systems.

The primary inputs to CASS include digital satellite data and US Geological Survey Digital Line Graph (DLG) data.

- The Landsat Thematic Mapper (TM) data provides 1:100 000 scale, 30-m resolution in seven bands. As in the manual procedure, TM data serve to identify strata according to land use and boundaries in the rural areas. SPOT panchromatic data with 10m resolution is used in highly urbanized areas.

- DLG data at a 1:100 000 scale are displayed as an overlay plane to aid in identifying physical boundaries on the TM data. They include the US Geological Survey's transportation and hydrography data.

- The US Bureau of the Census' digital map database (Topologically Integrated Geographic Encoding and Referencing System [TIGER]) is used to display political boundaries (state and county).

Although many procedures are automated in CASS, the stratification and formation of PSUs that require the identification of good physical boundaries is a visual process. Multitemporal and ground truth data needed for automatic classification based on the signatures corresponding to different land uses are very expensive.

TM bands two, three and four are used for optimal agricultural land use stratification. The digital nature of the data enables the analyst to create a colour map which best distinguishes cultivation and boundaries. Data from the latest June Agricultural Survey (JAS) provide ground truth. Colours are in three overlay planes (red, green and blue). The colouration is done by an iterative process, adjusting the intensity of each colour until the mix provides the best colour map for determining land use and identifying boundaries.

The 1:100 000 scale DLG boundary data are not updated regularly and 1:2 million scale DLG boundary data do not meet NASS' need for accuracy. Because the TM data are more recent than the DLG data, DLG data are used mainly to confirm boundaries that are questionable on the satellite data and are rarely used as the only source for a boundary. If the analyst can identify a physical boundary directly in the TM image, it is used. If a physical boundary cannot be distinguished in the TM image, but a DLG boundary exists, further research of other sources will be done before it is used. The TM image (scene) is precisely overlaid with DLG by registering the DLG data to the backdrop of satellite data. Several matching TM and DLG points are selected and a least-squares regression is run to fit the remainder of the data. These points and the regression are saved in a file and used each time a DLG file for that scene is displayed. The registration file also enables the user to determine latitude and longitude coordinates of any given point. The satellite image (usually a TM scene) is the base material. All other information in the overlay planes is registered to the image. Where there are differences, the image is assumed to be correct.

Other reference materials include 1:100 000 scale US Geological Survey maps, some small-scale aerial photography, topographic quadrangle maps for city areas, information from NASS State Offices and information on planting and harvesting dates for the major crops in the state. The SPOT satellite and the Indian Research Satellite (IRS) can furnish 10-m resolution for urban and other critical areas as needed.

2.3 Delineation of Primary Sampling Units and Strata

The construction of PSUs is completed county by county. The entire county must be covered with PSUs (no holes) and no PSU may overlap into an adjoining county. Political boundaries are loaded on one of the overlay planes. One of the first things that a person in charge of stratification must do is adjust the political county boundaries so that they fall on physical boundaries that can be located by an enumerator in the field. The new boundaries are filed and always appear when that image is loaded. State and national boundaries must remain in their political location.

The image analyst must consider which crops are grown in the state, what maturity phase the crops were in at the time the satellite image was taken and how to interpret the colours on the Landsat image.

The previous year's June Agricultural Survey data are used to aid the proper interpretation. The person in charge of the stratification photo interprets the colour Landsat image on the display screen and divides each piece of land into a PSU within a specific size range.

Under the CASS system, stratification and formation of PSUs are carried out simultaneously. Physical boundaries are identified on the Landsat image itself, enhanced by DLG data. This continues until all land areas in a county have been assigned to a stratum and a PSU. Quality checks are done to assure accuracy and consistency of stratification and boundary identification throughout the state.

In each county, polygons are drawn and tagged with a PSU number which consists of a stratum number and a sequence number. The particular stratum in which to place a unit of land is decided by interpreting land use as indicated by the colour TM image. At the same time, a PSU within some specific size range is delineated, using physical boundaries identified on the satellite image. In CASS, this is done by keying in a sequential PSU number provided by the software and then utilizing the mouse to select points along desired boundaries. When a PSU is closed, the area is immediately digitized and displayed. This allows the user to determine if the PSU is within the target size for that stratum. If polygons are too small or too large, they can be edited (combined, split, reshaped or deleted). When a county is completed, the analyst has the ability to check for overlapping polygons and holes (or missing land areas). At any time, a list of PSUs can be elaborated to check that PSU areas are within tolerance. The numbering created when the PSUs are being formed has no order as far as location is concerned and may skip numbers when PSUs are combined. When the stratification (formation of PSUs) is complete, the software will renumber the PSUs on command. In the reordering, the PSUs are numbered in a serpentine order within the county starting in the northeastern corner. The polygons are then saved to a file to be reviewed by another experienced analyst.

The ability to obtain images that correspond with certain crop growth stages and the use of JAS data for ground truth has led to the ability to identify and stratify accurately for some major crops such as cotton, rice, citrus, peanuts and wheat; even going so far as to separate winter wheat and spring wheat. State offices are a source of crop-specific information.

2.4 PSU Breakdown into Segments

After the entire state has been stratified and the total area for each stratum has been calculated, a separate program is run (outside CASS) to draw a sample of PSUs with probability proportional to size. These selected PSUs will be further divided into segments. Only those PSUs which were chosen by the sample selection program are divided into segments in CASS. The sample selection program retains three decimal places in the square mile areas of all PSUs. The user displays the STR (stratification) file saved (from the previous step) and enters in the PSU number to be sampled. Software then clears all but the sample PSU from the screen.

Each segment has a specific target size depending on its land use stratum. Each individual segment should have a size as close as possible to the target size using the best physical boundaries available. The person doing this work must also be able to interpret the satellite image to make each segment a good representation of the full PSU where land use (for instance, proportion of cultivated land) is concerned. Each PSU is then divided into a determined number of segments (say six to eight segments) depending on the stratum, as specified in Tables 4.2 to 4.6.

Many of the same functions which were involved in delineating PSUs during the stratification phase are used to divide the PSUs into equal sized segments. For example, the mouse is used to select points along an identifiable boundary. When the segment (polygon) is closed, the size is immediately displayed, and segments can be merged or split, or boundaries can be reshaped. Similar quality control checks for overlaps and omissions are done. Because segment areas are much smaller (typically one square mile), boundaries are often difficult to find. Occasionally, field edges, section lines or point to point must be used. When the PSU has been completely divided into segments, one is selected randomly using the segment selection command. Its latitude and longitude are then determined (for use in the NASS Geographic Information System) and a photo enlargement is ordered. Finally, the boundaries for sample segments are transferred to enlarged photos by the Sample Preparation Unit by displaying the segment or utilizing a print of the digital image and segment boundary. A map for each county showing the location of sample segments is also prepared.

2.5 Resource Considerations

The original purpose in developing CASS was to improve NASS area frames and use resources more efficiently. Stratification using all paper products was very labour-intensive. With CASS, digitization is built in and boundary transfers are eliminated, resulting in excellent labour saving (in the past labour was about 80% of the total cost, while with CASS the labour cost for the area frame development for Oklahoma was 42%).

On the other hand, material costs are significantly higher using CASS. The cost per scene of digital TM data is now approximately twice that of the paper product. For hardware and software resources, one-time and maintenance charges should be considered. NASA contributed to startup hardware expenses and software development. In the future, the system will need to be upgraded or replaced, but the present trend in workstation prices is decreasing while the amount of computing power per dollar is rising. At present, the cost of materials, hardware and software makes the cost of preparing a frame with CASS about the same as that of a paper frame.

2.6 Benefits

- The land use analysis is more accurate, as the scale of TM data has gone from 1:250 000 on paper to the 1:100 000 scale digital data. Furthermore, a dynamic colour map can be produced to enhance the image and bring out the cultivation, rather than having fixed colours as in the Landsat image print.
- The base display scale of the Landsat (1:100 000 currently) can be changed. Four different levels of magnification are used in the frame construction process. Most of the work is done in the highest level (level 4). The scale at each level is as follows:
  Level 1 = 1:100 000 (1/2 inch to a mile)
  Level 2 = 1:63 360 (1 inch to a mile)
  Level 3 = 1:42 240 (1 inch to a mile)
  Level 4 = 1:31 680 (2 inches to a mile)
- The automation of this process has eliminated the tedious, error-prone process of transferring boundaries from a satellite image print, to aerial photography, to a 1:100 000-scale USGS map, to a digital file. CASS also provides computer-generated random numbers for segment selection.
- The digital nature of data is a benefit. First, PSUs and segments can be more easily revised in CASS by moving boundaries, and because the size of the PSU or segment is known immediately and can be edited if it does not fall within the desired limits. A frame can be updated without doing a total recreation as would be necessary with paper frames. Backup copies of the digital data can be maintained off-site.
- Time and labour can be saved. Fewer people are required to complete the same number of states in a given period of time. For an average size State four persons are needed for delineation of PSUs and strata, and four for delineating segments.
- CASS makes it possible to maintain a latitude/longitude database of all selected segments, making it possible to use them as data layers in other Geographic Information Systems.

2.7 Concerns and Solutions

The identification of good physical boundaries is more difficult on TM than on aerial photography. Identifying potential boundaries may be facilitated by: i) using edge enhancing filters to make boundaries "stand out"; ii) using different bands or band combinations for boundaries; iii) using digital orthophotography (when available); and iv) future satellites may offer better resolution at a lower cost. Updated DLG data would also be a valuable help.

Other concerns are urban stratification and cloud cover. SPOT and IRS data with 10m resolution are being used for accurate urban frame construction and sample selection. When obtaining TM data, a compromise must sometimes occur between the most recent date and cloud cover. Cloud cover under 10% is accepted, although it depends on the cloud density and location. Aerial photography is ordered to supplement the TM in clouded areas.

During the development of new area frames using CASS for Oklahoma and California, the importance of proper registration of DLG to TM scene came out, particularly where TM scenes overlap. This problem has been resolved by adding sufficient control points.

Also, the Sampling Preparation Unit had a few problems transferring segment boundaries to photo enlargements. Problems were due to the age of photo enlargements and to the resolution of TM being lower than that of aerial photography.

The resource considerations of CASS (higher/lower costs) could change, depending on the future Landsat programme.

2.8 Two Pilot Tests for CASS

Once the initial CASS system was developed, two pilot projects were conducted in portions of the States of Missouri and Michigan. Enhancements were made as a result of these tests.

The initial test was conducted in Missouri in 1988-89 to gain experience with the software, to compare CASS with the paper-based method and to evaluate the speed of frame construction. Digital data covering three north-central counties of Missouri (Linn, Livingston and Macon) were used. These counties were chosen partly because the Area Frame Section of NASS had developed a new area frame for Missouri in 1987 for use in 1988.

The test proved that stratification using CASS was possible, reasonable and faster. Two weeks were required for paper stratification and digitization. CASS did the job in three days after set-up and problem resolution.

Five different people stratified each of the three counties. Two people used the paper method and three used CASS. The results illustrate the subjective nature of the work: for those using CASS, the percentage of the total three-county land area classified as >75% cultivated stratum, ranged from 23% to 37%; for those using the paper stratification ranged from 25% to 34%.

A more substantial test was done in Michigan, selecting a 21-county area for this purpose. At the time of the study (1989-1990), the state had just received a new frame in 1989 (implemented in 1990). Also, the Remote Sensing Section of NASS had recently completed work in the dry bean area of Michigan in regard to supervised classification; therefore, TM data was available. In this test, only one person worked each county since this was the way the system would be used operationally.

As before, a comparison was done between CASS and paper frame. The total area in each land use stratum was measured and the percentage difference between paper and CASS minus paper was calculated:

- in the stratum of >75% cultivated, the difference was only -4%;
- in the 50 to 74% cultivated, the difference was -14.6%, the analysts generally favoured the CASS results because they could better identify pasture, which is not considered cultivated;
- in the <15% cultivated stratum, a difference of +37.8% was found, the analysts again favoured CASS results because they felt they were better able to identify and include the woodland areas;
- in the agri-urban stratum, the difference of -14.3% was mostly due to the lower resolution of the TM data compared with aerial photography and quadrangle maps and, in this case, analysts favoured the paper result.

No ground truth verification was sought. Each of the five analysts visually compared the CASS and paper area frames for a given county, to see generally where differences occurred and why.

In general, because of the ability to change the displayed scale of digital TM versus the paper TM product previously used, and the dynamic rather than fixed colour mapping abilities, the CASS stratification was found to be more accurate than those developed from paper imagery. However, at that time, the stratification of urban areas was less accurate in CASS, because digital TM data (approximately 30m resolution, 1:100 000 scale) proved more difficult to interpret than the aerial photography (approximately 2m resolution, 1:40 000 scale) and quadrangle maps (1:24 000 scale) previously used. As already mentioned, SPOT and IRS data are now being used successfully for urban areas.

Table 4.1
United States: Area sampling frame sample sizes

State	June 1994	June 1995	June 1996	June 1997	Fall 1997	Frame year
Alabama	317	254	254	236	81	1991
Arizona	373	303	233	118	77	1984
Arkansas	395	395	395	348	176	1992
California	845	845	404	404	271	1994
Colorado	457	457	457	295	297	1984
Connecticut	47	47	29	8	11	1989
Delaware	72	72	58	23	18	1986
Florida	425	425	255	255	104	1983
Georgia	382	306	306	290	111	1991
Idaho	362	290	218	109	129	1982
Illinois	389	389	389	399	192	1975
Indiana	294	294	294	260	140	1976
Iowa	437	437	437	418	99	1989
Kansas	456	456	501	525	525	1996
Kentucky	338	338	338	179	156	1977
Louisiana	317	317	317	249	76	1990
Maine	117	117	71	32	21	1989
Maryland	252	202	152	61	44	1986
Massachusetts	67	67	41	12	15	1989
Michigan	312	312	345	145	121	1990
Minnesota	343	343	343	331	129	1975
Mississippi	402	402	402	441	124	1978
Missouri	387	387	387	335	236	1988
Montana	362	362	362	255	277	1986
Nebraska	390	390	390	443	259	1983
Nevada	104	84	46	26	18	1987
New Hampshire	47	47	29	10	8	1989
New Jersey	247	198	100	48	38	1987
New Mexico	292	292	292	154	109	1985
New York	380	320	256	256	76	1995
North Carolina	321	321	321	284	145	1978
North Dakota	376	376	376	420	78	1977
Ohio	289	289	250	220	174	1996
Oklahoma	405	405	405	353	361	1993
Oregon	372	372	372	163	246	1980
Pennsylvania	330	264	264	119	105	1981
Rhode Island	22	22	14	8	8	1989
South Carolina	335	270	270	119	71	1995
South Dakota	352	352	352	345	303	1976
Tennessee	382	349	349	306	170	1997
Texas	840	840	840	990	815	1982
Utah	324	261	198	69	62	1988
Vermont	62	62	38	21	21	1989
Virginia	343	260	260	146	118	1978
Washington	360	360	288	267	286	1980
West Virginia	165	132	66	66	35	1989
Wisconsin	310	310	310	207	127	1977
Wyoming	260	208	156	53	59	1985
TOTAL	**15 456**	**14 601**	**13 230**	**10 821**	**7 122**	

Source: USDA (May 1997). *June and Fall Area Frame Design Information for 1997.*

Table 4.2
Characteristics of the area sampling frame in the State of California (1994)

Stratum	Stratum definition	Square miles	Segment size	Number of segments in population	Number of June replicates	Number of June substrata	Segments in June sample	June Expansion Factor	Number of Fall replicates	Number of Fall substrata	Segments in Fall sample	Fall Expansion Factor
11	> 75% CULTIVATED	11 077	1.00	11 079	11	19	209	53	10	9	90	123
17	SPEC11: FRUIT, NUT	1 932	0.50	3 883	2	10	20	194	5	4	20	194
19	SPEC11: VEGETABLE	521	1.00	516	2	3	6	86	2	1	2	258
21	15-75% CULTIVATED	6 682	1.00	6 686	7	9	63	106	5	7	35	191
27	SPEC21: FRUIT, NUT	1 502	1.00	1 497	2	6	12	125	5	2	10	150
31	AGRI-URBAN: >100 HOME/SQMI	3 964	0.25	15 844	2	8	16	990	10	2	20	792
32	COMMERCIAL: >100 HOME/SQMI	901	0.10	9 021	2	1	2	4 511	2	1	2	4 511
41	<15% CULTIVATED	46 305	4.00	11 559	6	9	54	214	10	5	50	231
45	SPEC41: PUBLIC NO-AD, DESER	68 424	pps	2 470	20	1	20	124	40	1	40	62
50	NON-AGRICULTURAL	14 886	pps	396	2	1	2	198	2	1	2	198
		156 194		**62 951**			**404**				**271**	

Table 4.3
Characteristics of the area sampling frame in the State of Alabama (1991)

Stratum	Stratum definition	Square miles	Segment size	Number of segments in population	Number of June replicates	Number of June substrata	Segments in June sample	June expansion factor	Number of Fall replicates	Number of Fall substrata	Segments in Fall sample	Fall expansion factor
13	>50% CULTIVATED	4 191	1.00	4 219	6	13	78	54	5	3	15	281
20	15-50% CULTIVATED	10 328	1.00	10 406	10	9	90	116	5	5	25	416
31	AGRI-URBAN: >100 HOME/SQMI	642	0.25	2 575	2	2	4	644	2	1	2	1 288
32	COMMERCIAL: >100 HOME/SQMI	106	0.10	1 059	2	1	2	530	2	1	2	530
40	<15% CULTIVATED	35 333	2.00	17 737	6	10	60	296	5	7	35	507
50	NON-AGRICULTURAL	414	pps	43	2	1	2	22	2	1	2	22
		51 014		**36 039**			**236**				**81**	

Source: USDA (May, 1997). June and Fall Area Frame Design Information for 1997

Table 4.4
Characteristics of the area sampling frame in the State of Maine (1989)

Stratum	Stratum definition	Square miles	Segment size	Number of segments in population	Number of June replicates	Number of June substrata	Segments in June sample	June expansion factor	Number of Fall replicates	Number of Fall substrata	Segments in Fall sample	Fall expansion factor
14	25% OR MORE CULTIVATED	1 303	1.00	1 316	2	6	12	110	2	1	2	658
31	AGRI-URBAN: >100 HOME/SWMI	161	0.25	651	2	1	2	326	2	1	2	326
40	<25% CULTIVATED	29 351	4.00	7 346	2	8	16	459	5	3	15	490
50	NON-AGRICUTLURAL	473	pps	20	2	1	2	10	2	1	2	10
		31 288		9 333			32				21	

Table 4.5
Characteristics of the area sampling frame in the State of Oklahoma (1993)

Stratum	Stratum definition	Square miles	Segment size	Number of segments in population	Number of June replicates	Number of June substrata	Segments in June sample	June expansion factor	Number of Fall replicates	Number of Fall substrata	Segments in Fall sample	Fall expansion factor
11	>75% CULTIVATED	9 485	1.00	9 507	15	8	120	79	15	8	120	79
12	51-75% CULTIVATED	5 361	1.00	5 353	10	4	40	134	10	4	40	134
20	15-50% CULTIVATED	13 373	1.00	13 366	15	7	105	127	15	7	105	127
31	AGRI-URBAN: >100 HOME/SQMI	604	0.25	2 402	2	2	4	601	2	2	4	601
32	COMMERCIAL: >100 HOME./SQMI	115	0.10	1 153	2	1	2	577	2	1	2	577
40	<15% CULTIVATED	39 937	3.00	13 302	10	8	80	166	11	8	88	151
50	NON-AGRICULTURAL	192	pps	22	2	1	2	11	2	1	2	11
		69 067		45 105			353				361	

Source: USDA (May, 1997). June and Fall Area Frame Design Information for 1997.

Table 4.6
Characteristics of the area sampling frame in the State of Texas (1982)

Stratum	Stratum definition	Square miles	Segment size	Number of segments in population	Number of June replicates	Number of June substrata	Segments in June sample	June expansion factor	Number of Fall replicates	Number of Fall substrata	Segments in Fall sample	Fall expansion Factor
10	>50% CULTIVATED	21 241	1.00	21 411	15	15	225	95	15	11	165	130
14	SPEC13:COTTON	15 425	1.00	15 513	15	11	165	94	10	8	80	194
15	SPEC13:RICE	2 537	1.00	2 556	5	7	35	73	5	1	5	511
16	SPEC13:PEANUTS	919	1.00	929	5	2	10	93	5	1	5	186
17	SPEC13:WHEAT, SORGHUM	9 744	1.00	9 781	12	9	108	91	15	9	135	72
20	15-50% CULTIVATED	74 393	2.00	37 420	15	17	270	139	25	12	300	125
31	AGRI-URBAN:>20 HOME/SQMI	4 587	0.25	17 247	5	6	30	575	7	4	28	616
32	COMMERCIAL:>20 HOME/SQMI	3 262	0.10	32 622	5	7	35	932	6	3	18	1 812
34	SPEC13-20:POTENTIAL URBAN	1 029	0.25	4 119	5	4	20	206	2	1	2	2 060
41	SPEC40:OPEN RANGE	113 570	8.00	14 201	8	9	72	197	10	6	60	237
42	SPEC40:WOODS	13 344	4.00	3 344	5	3	15	223	5	3	15	223
50	NON-AGRICULTURAL	2 854	1.00	2 864	5	1	5	573	2	1	2	1 432
		262 905		**162 007**			**990**				815	

Source: USDA (May, 1997). June and Fall Area Frame Design Information for 199

Figure 4.1

CASS Workstation
The hardware setup with working screen and smaller command screen on the left

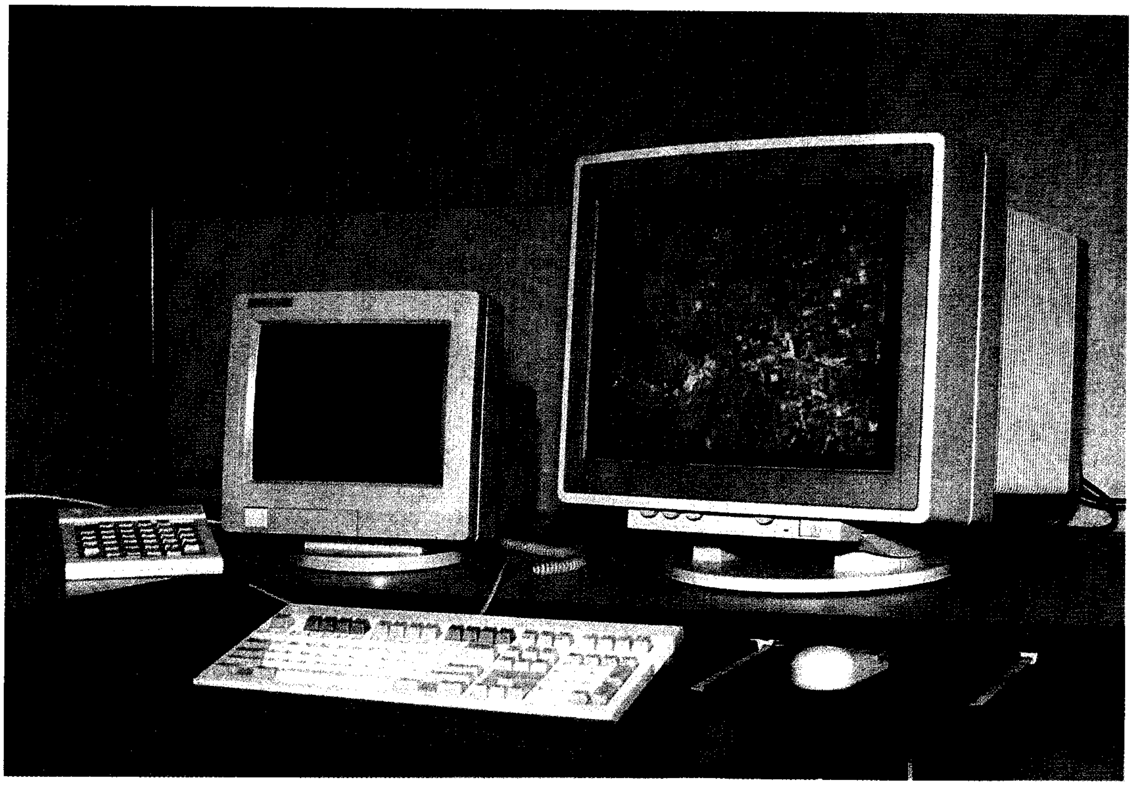


Figure 4.2

CASS Workstation
Illustration of the process of adjusting overlaid data to the image
The cross (+) indicates where the intersection (□) on the overplay plan should be located.
The software then makes the adjustment.

Figure 4.3

CASS Workstation
Illustration of the ability to change the colour of the overlay data for better visibility
(Compare with Fig. 4.2)

Figure 4.4

CASS Workstation
Screen with PSUs delineated

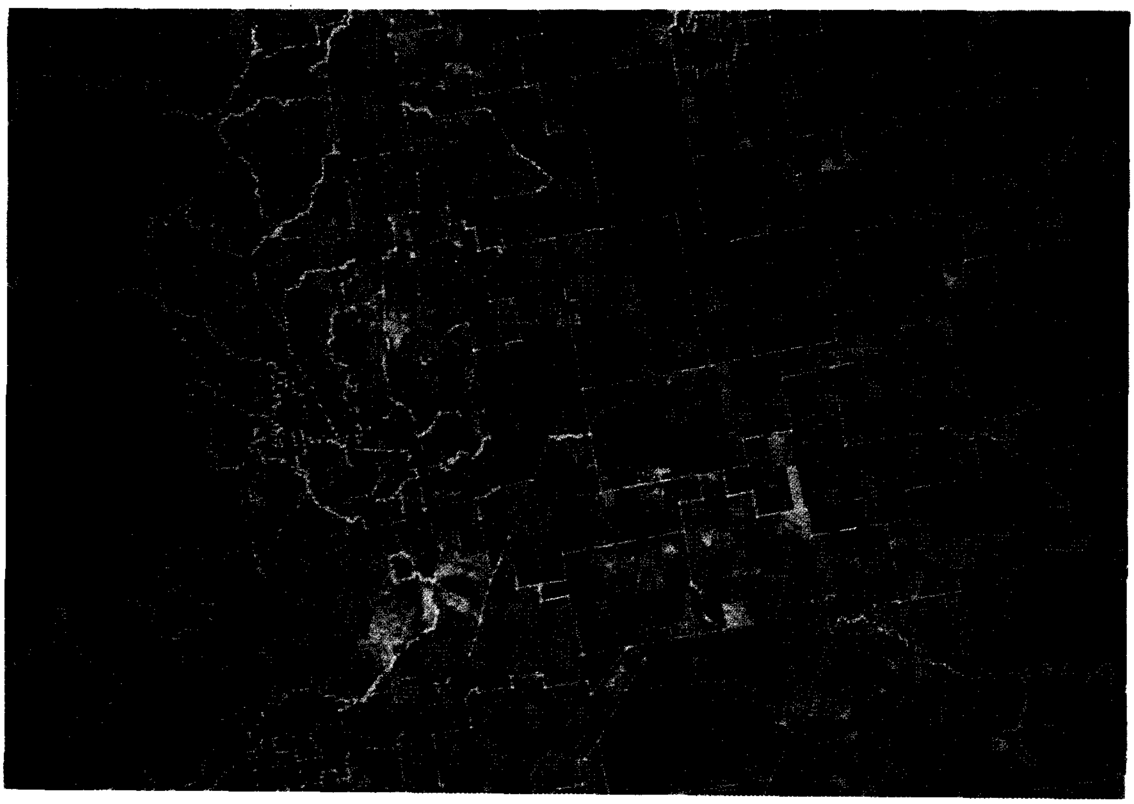


Figure 4.5

Screen with a selected PSU

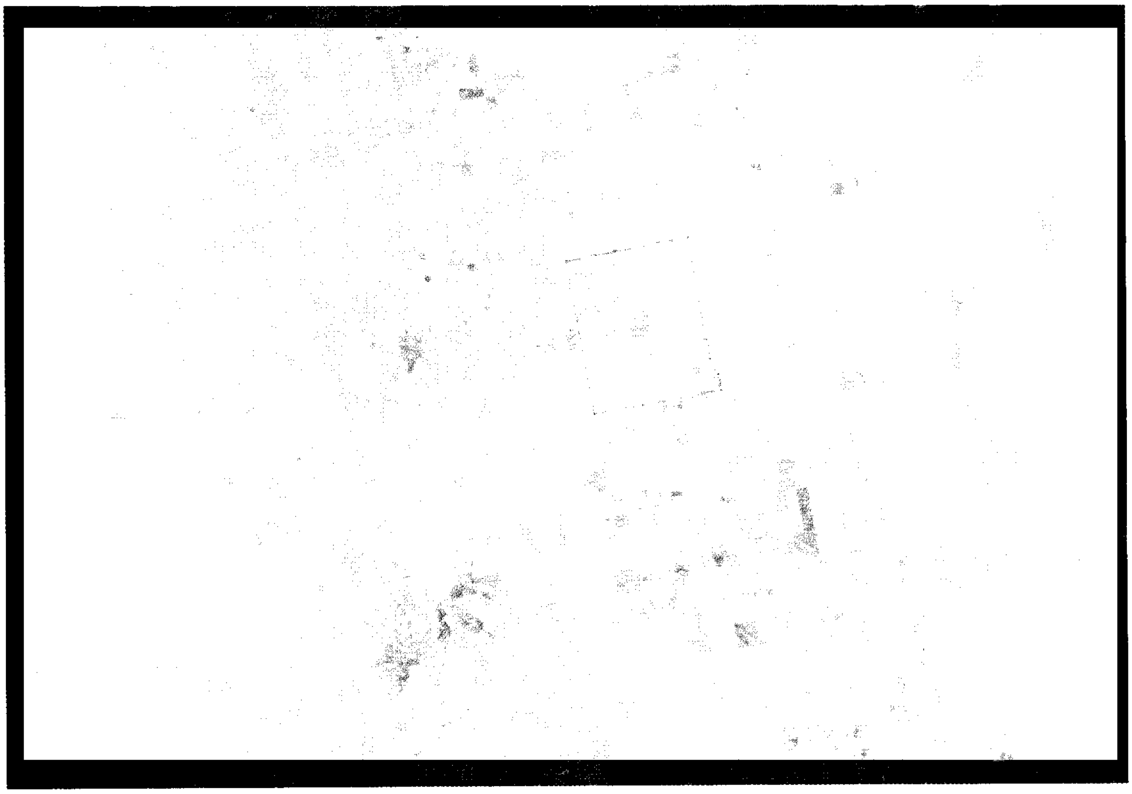
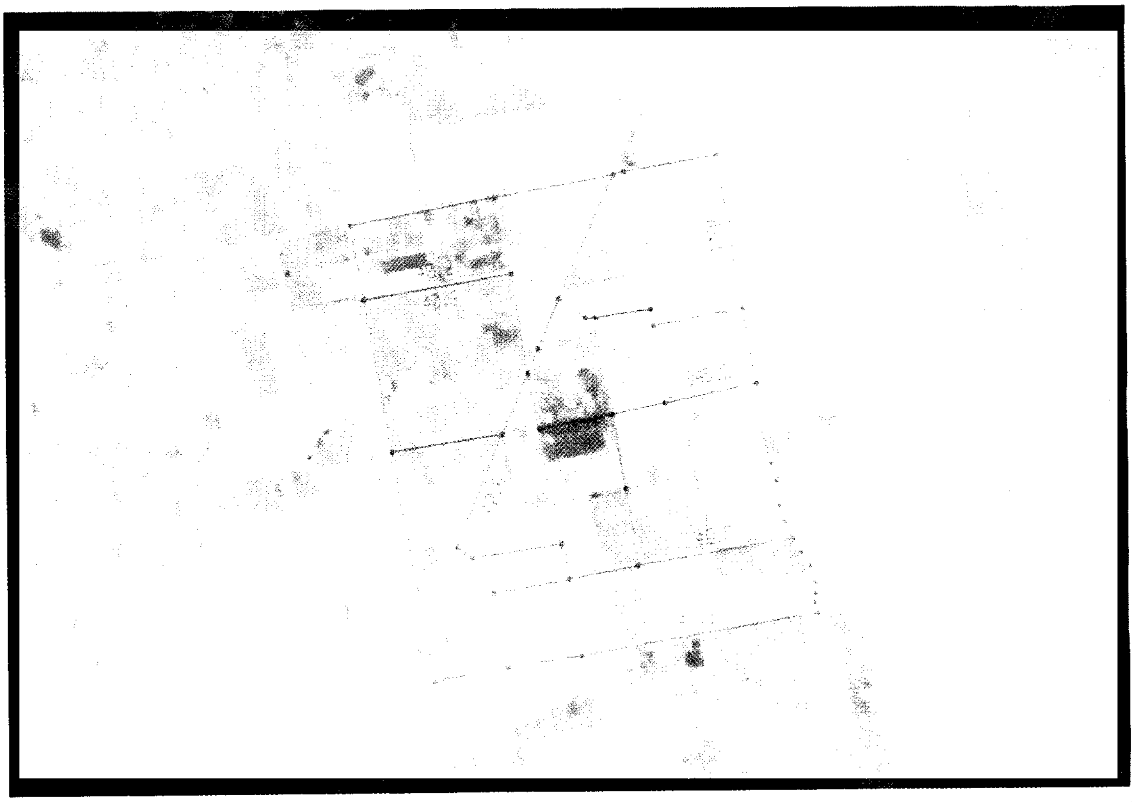


Figure 4.6

CASS Workstation
Screen with the PSU divided into Segments
(Segment #94502 represents a selected Segment)



CHAPTER 5

ALBANIA

GENERAL PURPOSE AGRICULTURAL SURVEY PROGRAMME BASED ON AREA FRAME SAMPLING METHODS (1992-1996)

1. INTRODUCTION

This chapter describes the construction and application of the area frame agricultural survey programme in Albania from its beginning in 1992 through the complete reconstruction of the frame to provide district estimates in 1996. The detailed description of frame construction covers the new frame (1996). So far, there have been no surveys carried out with the new sample. Very little information was available about survey activities in 1994 and 1995.

The agricultural sector is of primary importance in the Albanian economy. The recent change in Albania from a centrally planned economy to a market economy (March 1992), has affected the agricultural production system and caused short-term problems in maintaining food security and developing alternative distribution systems to replace centrally planned policies. Within the past few years the composition of the agricultural sector has changed from state and collective farms to an estimated 950 000 privately owned parcels, as determined by the 1993 Area Sample Survey.

The urgent need for agricultural information led to a request for technical assistance and the hasty construction of an area sampling frame and selection of an area sample in 1992. Two survey designs were used: for a large part of the country a survey design considered strata, primary sampling units (PSUs) and segments with recognizable physical boundaries, a design similar to the main survey design described, for example, in Volume I of this publication. For the remainder of the country, a survey design with square segments was used. The main objective of the May/June 1992 Area Sample Survey was to obtain national estimates for major crop areas, prospective wheat yields, and fertilizer use.

In 1993, the project was expanded to include the following goals: i) enhance the pilot area sampling frame through refinement of previous land use stratification in order to achieve greater homogeneity of land use within each stratum, thereby improving sampling efficiency; ii) train personnel from the Ministry of Agriculture and Food (MOAF) in conducting the area sample survey; iii) work in cooperation with the International Fertilizer Development Corporation (IFDC) and MOAF to conduct a crop area and fertilizer use survey (June 1993), using the refined area sampling frame; iv) use survey data to estimate Albania's 1993 planted crop areas; v) design a probability wheat crop cutting survey to generate estimates for Albania's 1993 wheat yield and production; and vi) establish a National Agricultural Statistics System (NASS) in Albania.

A number of improvements were introduced in the 1993 survey design. For instance, the part of the frame that consisted of square segments was discarded and all segments were defined with recognizable physical boundaries. Both the 1992 and 1993 samples were designed to provide estimates at the national level only.

In 1994, the Ministry of Agriculture and Food indicated the need of obtaining estimates at the prefecture (district) level. To meet this request, the decision was made to start with new materials and reconstruct the entire area frame. This would allow the correction of problems that had been identified when using the original area sampling frame. In 1996, a complete new frame was constructed using all new materials and the sample designed and selected to provide estimates for prefectures. Since then, political, economic and social instability have prevented the use of the new sample and made its future use uncertain.

2. GEOGRAPHIC AND POPULATION CHARACTERISTICS

Albania comprises an area of 28 740 km^2. As a mountainous country, intensive agriculture is contained along a coastal plain, extending into lowlands and valleys and up to foothills, uplands and mountains. Land elevation ranges from 200 m to above 1 000 m.

Approximately 80% of total land area in Albania is devoted to agriculture. Wheat, maize and vegetable crops are grown in the coastal regions and lowlands; tree crops are grown in the higher elevations. Winter wheat is the most important crop in Albania, with maize in second place.

Albania's total estimated population in 1992 was 3.3 million with a projection of 3.9 million for the year 2000. The population of the major Albanian cities is approximately 20% of the total population. In 1990, about 700 000 people were employed in agriculture and forestry, and an additional 100 000 in related activities.

3. RESUME OF MAIN METHODOLOGICAL FEATURES

The following general characteristics of the survey designs are contained in the 1992 pilot survey, the 1993 refined survey and in the new frame constructed in 1996. Features specific to each particular frame or survey are described in the following sections and subsections.

- Sample designs are based on probability sampling methods.
- The area sampling frames were prepared in order to:
  - provide complete coverage;
  - not be affected by changes in farm boundaries or changes of operators;
  - offer safeguards against omission and duplication; and
  - comprise sample segments that may be used more than once without replacement.
- Surveys are based on contacts with farmers operating in areas randomly selected from a constructed area sampling frame.

- Closed segment estimation methods are used for crops.

4. SAMPLE DESIGN

4.1 General Survey Characteristics

The 1992 area sample designs were used to conduct the crop, the fertilizer usage and the socio-economic farm-level surveys. The 1992 and 1993 surveys did not include livestock data, which will be collected in the future with the new area sample.

4.2 Stratification

Four strata were identified in the 1992 Pilot Survey: 1) intensive agriculture-along the coast and valleys;2) less intensive agriculture - mostly hill country; 3) cities and urban areas; and 4) lakes and open water. *Only the first two strata were included in* the survey.

Five strata were identified for the 1993 Survey: 1) coastal areas with intensive agriculture; 2) river valleys with intensive agriculture; 3) upland hills with diversified agriculture; 4A) low mountains with limited agriculture; and 4B) high mountains with little or no agriculture. A sixth stratum of non-agricultural land (i.e. lakes, cities and military complexes) was identified but not surveyed. The 1992 Pilot Survey Stratum 1 was not modified for the June 1993 Survey. The 1992 Stratum 2 was modified by dividing it into four, more specific strata (Strata 2, 3, 4A and 4B).

The 1996 stratification is based on proportions of the land being cultivated and does not take into consideration its location or slope (cf. Table 5.1 for a comparison of the three stratifications).

4.3 The 1992 and 1993 Area Sampling Frames

In 1992, two area sampling frames were used in the survey design for Albania. For Stratum 1, *intensive agriculture*, a traditional area frame (strata, PSUs and segments with identifiable physical boundaries) was constructed. For Stratum 2, *less intensive agriculture*, a grid-type area frame was constructed (refined in the 1993 Survey into Strata 2, 3, 4A and 4B). Strata 3 and 4 were not sampled.

In Stratum 1, stratification and formation of PSUs and segments were carried out on TM/Landsat images with a scale of 1:250 000. There were some boundary problems and it was difficult to identify segments. Once the segment was found, a drawing (not to scale) was made at the segment location to show the tracts and control the enumeration. With this process, the prime advantage of having a photographic enlargement to measure field areas was lost. Reliable objective measurement of agricultural areas could not be implemented. At the same time, there could be significant error in stratum and segment boundary identification. It is eminently better to have imagery at a scale of 1:100 000 plus photography at a scale of about 1:50 000 or smaller.

Grid Area Frame - Stratum 2. For the 1992 Pilot Survey, time constraints required the construction of a special area frame using MSS Landsat imagery, constructed by overlaying a grid over Stratum 2, formed by an equivalent to a 1 km^2grid on the ground. There were over 25 000 grid cells. A systematic sample of 40 grid cells (100 ha each) was selected by determining a sampling interval and a random start. After the random start, each 625th cell (25 000/40) was selected. Selected segments were transferred to 1:50 000 scale topographic maps. Grids were correlated to longitude and latitude in order to locate selected areas in the field using Global Positioning System (GPS) equipment. Each selected cell was divided into two equal parts. The locations were determined both by physical boundaries and by longitude and latitude using GPS. A systematic selection provides a sample that is arrayed by geographic location.

The majority of Stratum 2 segments were identified with little or no difficulty. However, some of the selected segments were found to be areas where there were no roads and/or no inhabitants. Personnel familiar with these remote rural areas identified segments which were inaccessible. For these areas, the crop survey questions were assigned a zero value on the assumption that there were no inhabitants and no crops.

The hills had tremendous diversity, including tree crops, and high mountains containing forest and non-agricultural land. These areas could not be stratified with MSS imagery. This area frame method was discarded in the 1993 Survey.

In 1993, the hill and mountain strata were separated and four strata were constructed, replacing the Stratum 2 used in 1992. With the division of Stratum 2 into Strata 2,3,4A and 4B, the traditional method of frame construction and sample selection was applied for the whole country. Table 5.2 presents some survey parameters used in the 1993 Survey.

4.4 The 1996 Area Frame Construction

The new frame was constructed for each prefecture using the following steps:

Step 1. Strata identification on maps;
Step 2. Dividing strata into PSUs;
Step 3. Numbering the PSUs;
Step 4. Digitizing and processing the information on the computer;
Step 5. Selecting the sample PSUs for later splitting into segments.

Step 1. Strata identification for each prefecture was based on the following procedure:

- Entering the prefecture boundaries from the 1:100 000 scale topographical maps on transparency paper and overlaying them on the satellite images to locate the prefectures.
- Careful observation of satellite images (free observation or by magnifying glass) to define the zones of different strata and identify the physical boundaries to be delineated on the transparency paper.

- Laying the transparency paper over the topographical maps and careful observation of them to coordinate the information obtained from the satellite images especially related to the physical boundaries.Next, strata boundaries were entered on these maps. This procedure was done with the help of prefecture specialists who knew the areas and provided assistance in taking decisions in disputable cases.

Step 2. Dividing the strata into PSUs. Based on previous experience with the old frame and the results of completed surveys, the size of PSUs by stratum was defined as follows:

Stratum	Size of PSUs		
	Minimum	Target	Maximum
1	800 ha	1 000 ha	1 200 ha
2	8 00 ha	1 000 ha	1 200 ha
3	1 600 ha	2 000 ha	2 400 ha
4	1 600 ha	2 000 ha	2 400 ha

Procedure:

- The areas of the strata were first entered on 1:100 000 scale transparencies.
- PSUs were then formed on the transparencies using the size definitions from the above table based on identifiable boundaries on the topographic maps. The areas were measured using a grid (1 square = 100 ha).
- The transparencies were then laid over the satellite image to verify two important elements: uniformity and existence of identifying boundaries.
- The transparencies were then put back over the topographical maps and PSU boundaries and were then entered on the maps using permanent yellow colour.

It should be noted that in some cases there were significant deviations from the size of PSU defined in the above table because of the lack of visible physical boundaries, especially in Strata 3 and 4. For the zones where it was impossible to find physical boundaries imaginary boundaries were used.

Step 3. Numbering of PSUs was completed by prefecture, and within each prefecture by strata. The following procedure of numbering was used: numbering started from the northeastern corner of each stratum. The PSUs were numbered in a serpentine manner. The PSU identification number includes the prefecture number (by alphabetical order), the stratum number and the PSU number. Finally, the total number of PSUs was calculated.

Step 4. Computer digitizing and processing of PSU information was done using a geographic information system (GIS). The basic programs used to digitize and process the data were PC/ARCINFO Ver. # 3.41 and ARCVIEW 2.1. Briefly, the steps followed to process the information were:

- Prefecture and PSU boundaries were digitized from the 1:100 000 topographical maps using ARCEDIT.
- The data were edited through ARCEDIT.
- Transformation of the data into real coordinates was done using the modules designed by SML program and DBase IV (version of ARCINFO executed in DOS).
- Data analyses to calculate PSU areas to be used later in selecting the sample PSUs and performing other statistical analyses were completed.
- Generating information in a mapping form with different scale was completed using the ARCPLOT module and other procedures designed in SML and ARCVIEW 2.1. In addition to the mapping information, tables summarizing the complete information were generated for each prefecture.

There is now a database related to the designed frame, which can be used in the future when making decisions on frame adjustments.

4.5 Sample Selection

4.5.1 Sample Segment Identification on Maps and on the Ground

The total segment sampling size was 400 segments based on an assessment of the capabilities of MOAF to collect and process quality data, and provide estimates of the required precision. This is a 37% increase in sample size over the previous area sample.

Once this total sample size had been determined, the first step in allocating the sample to strata was the calculation of the standard errors by strata for a representative number of crops (wheat, maize, white beans and alfalfa) and livestock (cattle, goats, sheep and pigs) for 1994 and 1995. The second step was the allocation of the 400 sample segments between strata using the Neyman optimum allocation procedure. The last step was the allocation of segments by strata among prefectures (defining the number of segments for each prefecture). The sample segments were then summed across the prefectures to ensure a minimum sample of 20-30 segments at the prefecture level.

4.5.2 Segment Selection

A PSU worksheet was prepared for each prefecture listing each PSU, its area (ha), the calculated number of segments for each PSU (knowing that the desired segment size for Stratum 1 = 25 ha, Stratum 2 = 50 ha, and Strata 3 and 4 = 100 ha), and the accumulated number of segments for each stratum. It was estimated that there would be 20 farm interviews per segment of approximately 25 ha.

The sample PSUs were defined using systematic selection with a random start. The sampling interval was defined by dividing the accumulated segment numbers for each stratum (total segments) by the stratum number of sample segments. Sample PSUs were then defined as those which included the segments selected systematically. Then, the selected PSU was split into two approximately equal parts (A and B). Next, it was determined whether the selected random start number fell in part A or part B. Since the selection of PSUs was based on a cumulative column of potential segments, it would be immediately known whether the A or B half contained the selected segment. However, the selected half was alternated so that the selection was not allowed to fall twice in a row within the same half. The plastic grid was then placed over the selected half so that the assigned number of segments could be made to fit in that half. Next came the serpentine numbering on the plastic grid. After numbering, the sample segment was selected at random.

This procedure was followed for all PSUs and segments within a prefecture for all prefectures. The sample PSUs shown on the 1:100 000 scale maps were then transferred to 1:25 000 scale maps (covering all the country) and 1:10 000 scale maps (where available). See Table 5.3 for the sample allocation.

Segment identification was as follows:

- The selected PSU was split into two approximately equal parts. If the orientation of the PSU was in a vertical plane (north to south), the northern portion became Part A and the southern portion Part B. If the orientation of the PSU polygon was in a horizontal plane (east-west), the eastern portion became Part A and the western portion Part B.
- The part of the PSU where the selected segment lay was entered on 1:10 000 (where available) and 1:25 000 maps.
- Grids were constructed corresponding to the desired segment sizes and map scales used.
- Serpentine numbering of the grids beginning in the northeastern corner was completed for each portion of a PSU.
- Using the selected grid unit which corresponded to the sample segment, boundaries were located to construct a segment with a size of plus or minus 10% of the desired segment size for the stratum. The following rules were established: the constructed real boundaries of the segment should contain a portion of the grid square or at least touch the grid boundary; stratum boundaries could not be crossed; crossing into another PSU was allowed as long as a portion of the segment was in the selected PSU; if boundaries of the segment crossed a prefecture boundary, more than 50% of the segment should be in the prefecture from which it was selected; it was preferred not to cross prefecture boundaries. When this happened, a relocation of the segment boundaries should be done. If this could not be accomplished, then an adjoining grid segment within the PSU should be selected. After final review, each selected segment was measured on the digitizer to assure that the size was within the tolerance limits.
- Ground segment location was done using 1:10 000 or 1:25 000 maps and, in some special cases, Global Positioning Systems (GPS). Central and district personnel assisted in this process. After it was defined whether the selected

segments contained farming activities or dwelling units, a complete list of farmers was prepared which was used as the secondary sample.

Digitizing and processing of segment information was done using the same procedure as with PSUs.

4.6 Data Expansion and Calculation of Sampling Errors

In the 1992 Pilot Survey direct expansion was applied to Stratum 1 and ratio estimation to Stratum 2. A significant amount of wheat and maize was grown on State farms along the coastal plain, identified as Stratum 1. A list of all State farms, obtained from MOAF, was used to obtain multiple frame estimates. In 1993, as most of the land formerly belonging to State farms had been privatized, multiple frame estimators were no longer necessary.

A direct expansion estimate will be used with the new area sample. Direct expansions will be used in future surveys coupled with a ratio estimator to measure change from one survey to the next.

The formula applied to calculate sampling errors for the area sample can be seen in Chapter 11. Systematic sampling was not considered in the formula.

5. SURVEY DATA COLLECTION (1992 and 1993 SURVEYS)

5.1 Selection and Training of Field Personnel

The focus for the 1992 and 1993 surveys was on area of major crops, production of winter wheat and fertilizer use. In the 1992 Pilot Survey, 40 enumerators and 11 English-speaking supervisors were selected. Training materials developed in English had not been translated into Albanian. Half of the enumerators performed the crop survey and the other half the socio-economic survey. After an intensive three-day training session (two-day classroom; one-day field trial), the group was divided into ten teams, each team consisting of four enumerators and one supervisor. Each team was assigned a car and driver. In the 1993 refined survey, economists and agronomists of the Ministry of Agriculture and Food, located in the vicinity of selected segments, were trained in survey data collection procedures. The number of sample segments was increased from 100 segments in 1992, to 210 segments in 1993. Consequently, 420 enumerators - two per segment, one economist and one agronomist - were trained to conduct the survey.

No survey has been conducted using the new sample of 400 segments which would double the workload of the previous survey.

5.2 Organization of Fieldwork

Sample segments were marked on maps and satellite images, enlarged to a scale of 1:50 000. Each team was assigned segments and provided with appropriate copies of these materials. Fieldwork began by i) driving to the segment(s), ii) identifying the boundaries of each segment on satellite images and/or aerial photos and photocopied maps, iii) locating the farmers with farm inside the segments, and iv) collecting data for the land inside each segment.

During data collection a sketch of the land was made. This served as a data quality check to assure that all segments were surveyed. Both the 1992 and 1993 surveys contained the same crop categories with the addition of one category, "Other Crops", in the 1993 Survey. The Wheat category was "wheat planted" for the 1992 Survey, and "wheat harvested" for the 1993 Survey. Also, in the 1993 Survey, "Maize" was displayed in two categories, "for grain", and "for silage".

5.3 Quality Control

During data collection, a 10% sample of the segments in Stratum 1 was selected for a reliability check. Supervisory staff went to the selected segments, talked with previously interviewed farmers and rechecked the field sizes. After verifying that i) survey teams had located segments correctly, ii) farmers interviewed had provided correct information, and iii) farmers' names were correct, it was determined that the interviewer's results were satisfactory.

6. DATA PROCESSING (1992 and 1993 SURVEYS)

6.1 Data Editing

Manual field edits were performed by a supervisor during data collection. Before data entry, questionnaires were reviewed by enumerators and supervisory staff. If additional information was needed or some items were questionable, the enumerator returned to the field to verify the data. Computer edits were performed during data entry and after data entry was completed, to obtain "clean" data sets.

6.2 Data Entry/Computer Programme

More than 100 data entry and edit programmes written in DBase III were used for data entry. These DBase III programmes are based on programmes developed by the Monitoring Agriculture with Remote Sensing (MARS) project at the Joint Research Centre of the European Union. Data were entered into two PC computers with 120 megabyte hard disk drives using a DBase III programme modified to fit the Albanian questionnaire. The programme is designed to display the questionnaire form one page at a time. As computer operators enter data from the completed questionnaire, many significant checks incorporated in the programme are performed. Other functions incorporated in the DBase III programme include:

- At the conclusion of data entry for each segment, all land areas contained in the questionnaires pertaining to the segment are summed up. This provides a check of segment completeness. Should the result of the sum not equal the total segment size, data must be rechecked to find the omission(s).
- Total number of hectares for each crop and amounts of fertilizer are indicated.

- Summarization of all items surveyed for each segment is performed. This enables calculation of the expanded total for the strata.
- Calculation of expansion factors is performed and expanded data and estimates are presented in table format.

7. DATA ANALYSIS AND RESULTS (1992 and 1993 SURVEYS)

The 1992 Pilot Survey was the first national survey conducted in Albania, after nearly 50 years under a planned economy. During this time, Albanian crops were raised on land operated as State or cooperative farms, and crop production information was based on data extracted from farm reports and sent to the capital. These data were summarized as by-products of administrative records. No sampling methods were applied.

The 1993 Survey provides a comparative database for future surveys. As a pilot survey, the 1992 Survey used a different sampling frame and a much smaller sample than (the refined) 1993 Survey. A significant difference is revealed by comparing the 1992 Survey data with 1993 Survey data.

7.1 Results of 1992 Survey

The area sample survey was designed to: i) determine the area of each major crop to be harvested in Albania in 1992; ii) prepare a forecast of winter wheat production in 1992; and iii) determine the kinds and amounts of fertilizer applied.

The survey was conducted during the period 15 May-9 June 1992 prior to crop harvests. See Table 5.4 for June 1992 Survey data on "Cropland Area by Land Use and Stratum".

***ASF data* vis-à-vis *MOAF data*.** The area sample measurements determined 2 768 600 ha of total land. MOAF data indicated 2 874 800 ha of total land. The 3.7% difference might be attributed to: i) the fact that area under water was not measured in the ASF; ii) an accumulation of measuring errors in Stratum 1 owing to small biases associated with grid measurement; iii) the fact that land measured on maps and satellite images does not account for all land area owing to slopes that are not projected correctly; or iv) the lack of precision of MOAF data.

***Fertilizer usage*.** Another purpose of the 1992 Survey was to estimate the amount of fertilizer usage. Consequently, the questionnaire contained a set of questions to elicit information as to whether fertilizer had been applied and, if so, what kind and in what quantity. The data indicated that very little fertilizer had been applied at the time of the survey in May and June. The IFDC/USAID fertilizer auctions were in progress at the time of the survey. The acquired data were of such a low percentage that it was not possible to expand the data and obtain reliable estimates of fertilizers applied. The survey timing did not allow for collection of data on fertilizer usage for the summer crops planted about the time of the survey. Moreover, the IFDC fertilizer import programme could have no impact on the winter small grains because of the late start of the programme.

Idle land. Estimates of idle land seemed high compared with other years. Factors which may have contributed to the increase of idle land in 1992, as reported by farmers, may be that: i) some land had not been planted because the boundaries of private land were not clear; ii) land had not been distributed in time to be planted and was still to be distributed; iii) in some cases, farmers reported that seed was not available, or farmers did not know how to proceed. Furthermore, during the review process, it was noted that some land in mountainous areas had been identified as idle which should be classified as wasteland. For future surveys, it will be necessary to follow stricter survey procedures.

7.2 Results of the 1993 NASS Survey

Survey questionnaires for Albania's 1993 National Agricultural Statistics System (NASS) Survey were nearly identical to the 1992 questionnaires. In both surveys, questionnaires were developed to obtain crop areas and fertilizer usage. MOAF economists and agronomists, located in the vicinity of selected segments, were trained in survey data collection methods. Preliminary data evaluations indicate that the enumerators collected good-quality data. Table 5.6 shows summary results of the survey.

7.3 Results of the 1993 Wheat Yield Survey

Yield survey methods

A subjective estimate of the wheat yield was obtained by asking farmers what yield they expected to obtain from each of the surveyed fields. A sample of fields for crop cutting (Objective Yield Survey) was also selected to provide training in the procedure and to improve the overall estimate. Since Albania's wheat harvest starts in early June and the NASS Survey was scheduled to be completed in Stratum 1 on 10 June, it was decided to select wheat sample fields for the Objective Yield Survey in this stratum as data were received and processed. This procedure was used only in Stratum 1. The fields were selected with probability proportional to size to select representative fields. The order of the segments selected was in the order of how the segments were completed and processed.

- Fifty samples in Stratum 1 were selected from the 60 segments of the 1993 sample. Fields on farms reporting wheat for grain harvest had a chance of being selected for preharvest wheat cuts in proportion to the number of hectares in the segment. There were 47 samples processed for the average yield data in this stratum.
- All 40 segments of Stratum 2 were selected. Enumerators were instructed to take the sample in the first wheat field they found in the segment. (A better procedure would have been to select an individual field and have the enumerator return to the exact field.) The team was aware that there would be some segments without wheat fields. Twenty-five samples were processed in Stratum 2.

- Thirty-five segments in Stratum 3 and ten segments in Stratum 4A were selected. Enumerators were again instructed to take the sample in the first wheat field found in the segment. Nine preharvest samples were obtained for Stratum 3, and one in Stratum 4A.
- No samples were allocated to Stratum 4B (high mountain region).

The complete survey included 82 fields of two plots each, or 164 samples. The sample unit was 0.5 m^2. The wheat samples were processed for total yield and moisture percentage, and adjusted to a 12% moisture level. A gleanings survey indicated that during harvest there was an average loss of 220 kg per hectare. This was subtracted from the average yield calculated from the crop cut sample data in each stratum.

Yield data displayed in Table 5.7 are averages calculated using the farmers' expected yields as reported on the NASS Survey, and the yields calculated from the crop cut survey. Each stratum's average yield is multiplied by the stratum's wheat hectares harvested, also reported for the NASS Survey, to obtain total wheat crop production for Albania. In the future, the wheat yield survey will be run using self-weighting samples.

7.4 <u>Results of the Fertilizer Use Survey</u>

Each farmer surveyed for the 1993 NASS Survey provided information on the type(s) and quantities of fertilizer applied to each field in 1993. Use of fertilizer is increasing as distribution systems are put into place. Use of organic fertilizer predominates over chemical fertilizers; quantities (metric tons) used for five fertilizer categories are displayed in Table 5.5.

8. CONCLUSIONS [1]

*In order to maintain an agricultural statistics system which supports the decision-*makers of both the Ministry of Agriculture and Food (MOAF) and the private sector, it is recommended that the National Agricultural Statistics System (NASS) contain a plan to:

- Continue national-level surveys not only for crops, but also for livestock, and other important social and economic items. Personnel training and questionnaire design would be required.
- Involve key decision-makers within the agricultural sector very early in the process. They should have great influence in determining the total budget available. Both current and potential survey data users in the public and private sectors should determine their relevant data needs.
- Establish within MOAF an advisory group of experts in data analysis. This group should be empowered to evaluate data and provide decision-makers with projected outcomes relevant to current, alternative and possible future decisions based on their analysis.

---

[1] Prepared by the statisticians of the entity providing technical assistance for the 1992 and 1993 Surveys.

- Have the Division of Statistics, MOAF, develop sampling procedures designed to obtain estimates at the prefecture level. This would by necessity increase the sample size.
- Improve segment identification materials. Prior to the next NASS survey, each segment selected for use should be remeasured in hectares. Acquisition of aerial photographs would be ideal. Use of aerial photography in the field would assist in the identification of segments by both the enumerators and the farmers. The recalculated measurements should be used by enumerators as a control during data collection.
- Designate one agency within MOAF to manage the area frame master materials. It would have responsibility for quality control and maintainance of the physical integrity of all materials.
- Have a survey quality check performed by the central staff immediately after a NASS survey is completed. This would entail re-interviewing a 5 % subsample of the segments surveyed.
- Run the NASS survey in May. Data processing and publication should be completed before 5 June of each year. This earlier completion date would make it possible to provide a PPS sample of fields and a self-weighted yield estimate for a wheat objective yield survey.
- Involve MOAF professional staff as much as possible in the development of NASS to assure quality data and facilitate training. As survey items are added, experts in each area of the survey should be employed whenever possible.

9. COMMENTARY ON 1995 SURVEY PROCEDURES [2]

Agriculture in Albania is characterized by many small farms. Data from the 1994 screening survey show that approximately 60% of the segments need to be reduced in size if it is assumed that 14 or fewer tracts represent an acceptable number. Enumeration workload could be reduced by subdividing the segments if aerial photography and sufficient identifiable boundaries were available.Lacking this, screening surveys are carried out in which all farmers having land inside the segment boundaries are listed. From the list, six names are selected for enumeration.

Screening is a difficult task that adds nothing to the overall survey information. The sample segments need to be screened every year. In addition, screening is prone to undercoverage. Reasons for undercoverage include not going into segments located a long way from the road, not making a sufficient number of callbacks and/or becoming confused and inadvertently missing a portion of the segment. Every effort should be made to utilize whatever photography and maps are available to subdivide segments with an excessive number of tracts.

The screening forms were sent to the enumerators *without* a definition for a farm or a farm operator. This gives rise to non-sampling errors since there are probably as many definitions of a farm as there are enumerators. Also, it appears that it was left to the enumerator's ingenuity to devise the operational procedure for developing a list of all

[2] See Beller, 1995.

farms associated with a segment. The listing is a large job and some parcels could easily have been missed, thus introducing a downward bias. The screening survey can be improved by providing the needed definitions and by having questions on the screening form that include the key points of the definitions such as "Do you or anyone else in the household have land that you worked on or any livestock?". Additional questions should further determine the existence, or not, of a farm and/or a farm operator.

In Albania,nearly 30% of the segments had farm operators with expansion factors greater than 1 000.For example, take a segment that has an expansion factor of 1 200. If 1.9 ha on the farm is rounded to 2 ha, the change in the estimate is 120 ha. The weights for the weighted estimator are calculated by dividing the total land operated inside the segment by the total land in the entire farm. Data collected on the entire farm are prorated into the segment by multiplying the entire farm information by the weight. The lack of provisions for handling the land within the segment that is not planted to any crop is a possible source of non-sampling error. Within a parcel there may be land devoted to a small clump of trees, a rocky area, irrigation ditches, fencelines, other wasteland, etc. Bias occurs when these non-crop areas exist and only net crop area is reported. Conversely, bias also occurs when the waste areas are included in the reported crop area.

The enumerator's manual does not stress or even mention that the enumerator should have the respondent accompany him to the segment during the interview. It is possible that an incorrect parcel is placed in the segment unless the interview is consistently conducted in the segment. Non-sampling errors are introduced when this happens. It is also possible that a portion of the parcel is reported as being entirely and not only partially in the segment, thus causing a serious error in the farm weight. Finally, if the interview is always held in the segment, it is easier for the enumerator to assure whether the area not devoted to cropping in the parcel was included or not. The manuals and questionnaires do not address this issue.

Table 5.1

Albania: land use strata definitions

1992 Strata	Description
Stratum 1	With intensive agriculture along the coast and valleys
Stratum 2	Primarily hill country with less intensive agriculture
1993 Strata	**Description**
Stratum 1	Coastal areas with intensive agriculture
Stratum 2	River valleys with intensive agriculture
Stratum 3	Upland hills with diversified agriculture
Stratum 4A	Low mountains with limited agriculture
Stratum 4B	High mountains with little or no agriculture
1996 Strata	**Description**
Stratum 1	Zones with more than 75% cultivation
Stratum 2	Zones with 25% to 74.9% cultivation
Stratum 3	Zones with less than 25% cultivation
Stratum 4	Zones with essentially no cultivation
Stratum 5	Military zones
Stratum 6	Main cities
Stratum 7	Water - large rivers and lakes

Only the first four strata were surveyed.

Table 5.2

Albania: 1993 NASS Survey parameters

Stratum	Land area (% of total)	Primary sampling units (PSUs)	Sample units (segments)	Target sample unit size (ha)	Sample units selected
1	11	293	10 902	25	60
2	14	385	13 684	25	40
3	43	601	21 539	50	70
4A	10	87	2 474	100	20
4B	19	185	4 825	100	20
Total	**97**	**1 551**	**53 427**		**210**

Table 5.3
Albania, 1996: total and sample segments

Prefecture	Stratum								Total sample
	I		II		III		IV		
	N	n	N	n	N	n	N	n	n
Berat	570	5	361	5	951	12	551	3	25
Durres	1 598	13	351	5	187	3	-	-	21
Dibra	-	-	653	9	1 113	14	1 011	4	27
Elbasan	626	6	636	7	2 087	27	720	3	43
Fier	3 681	31	799	11	394	5	104	2	49
Gjirokaster	582	5	-	-	1 575	20	1 124	4	29
Kukes	-	-	308	8	772	10	1 378	5	23
Lezha	1 242	10	-	-	968	12	430	2	24
Tirana	1 016	9	790	11	925	12	-	-	32
Vlora	1 420	12	1 016	10	990	13	781	3	38
Shkodra	1 943	16	242	3	1 978	25	740	3	47
Korça	1 665	14	247	3	1 603	20	1 403	5	42
Total	**14 338**	**121**	**5 399**	**72**	**13 562**	**173**	**8 245**	**34**	**400**

Table 5.4

Albania, June 1992 Survey: cropland area by land use and stratum.

Land use	National level (ha)	Stratum 1 (ha)	Stratum 2 (ha)
Winter wheat planted	132 200	62 200	70 000
Maize planted	76 900	21 100	55 800
Alfalfa	47 000	24 400	22 600
Vegetables and melons	35 700	21 500	14 200
Tree Fruit	25 800	9 700	16 100
Olives	19 600	1 700	17 900
Dry beans	15 600	8 700	6 900
Tobacco	14 600	4 000	10 600
Vineyards	8 200	6 000	2 200
Sunflower	8 000	8 000	0
Fodder	5 100	2 500	2 600
Potatoes	4 700	1 400	3 300
Cotton	2 200	2 200	0
Soybeans	1 800	0	1 800
Barley	1 500	800	700
Sugar beets	1 300	900	400
Oats	700	300	400
Total cropland	400 900	175 400	225 500
Idle cropland	187 600	48 100	139 500
Total crop and idle	588 500	223 500	365 000
Pasture, forest wasteland and non-agricultural	2 180 100	29 600	2 150 500
Total land Albania	**2 768 600**	**253 100**	**2 515 500**

Table 5.5

Albania -1993: fertilizer use by stratum

(metric tonnes)

Fertilizer	Strata					
	1	2	3	4A	4B	Total
Urea	11 447	10 044	8 939	1 012	207	31 649
Ammonium nitrate	7 152	1 505	5 083	327	169	14 236
Single super-phosphate	2 093	2 627	1 529	0	0	6 249
Diammonium phosphate	3 925	342	388	57	0	4 712
Organic	248 566	205 647	454 666	113 173	62 146	1 084 198

Table 5.6
Albania
NASS 1993 June Survey: land use by stratum

Crop	Strata (ha)					
	1	2	3	4A	4B	Total
Winter wheat harvested[1]	76 800	38 200	36 600	39 000	0	155 500
Maize for grain	37 800	23 000	33 600	7 400	2 000	103 800
Maize for silage	2 600	2 600	3 900	- - -	300	9 400
Alfalfa	27 900	5 700	9 700	1 400	500	45 200
Tree fruit	14 900	8 800	6 700	14 500	100	45 000
Vineyards	9 700	4 900	2 800	0	600	18 000
Dry beans	9 600	3 000	4 700	800	0	18 100
Vegetables and melons	8 500	4 700	3 400	700	500	17 800
Fodder[2]	4 900	500	3 900	1 100	0	10 400
Tobacco	4 600	5 100	1 700	0	0	11 400
Potatoes	2 100	1 600	5 800	600	200	10 300
Sugar beets	1 600	0	0	0	0	1 600
Cotton	1 500	0	0	0	0	1 500
Olives	1 500	44 100	6 700	0	0	52 300
Barley	1 400	100	200	0	0	1 700
Oats	1 300	1 400	3 000	1 200	0	6 900
Sunflower	300	400	0	0	0	700
Soybeans	100	0	0	0	0	100
Other crops	5 000	700	4 700	2 000	700	13 100
Idle cropland	43 100	20 900	28 600	7 000	2 300	101 900
Pasture, forest and waste	49 335	205 412	1 064 508	270 900	547 500	2 137 655
Total	**304 535**	**371 112**	**1 220 508**	**311 500**	**554 700**	**2 762 355**
Cities						9 049
Military						26 735
Lakes						45 622
Measurement bias[3]						31 239
Grand Total						**2 875 000**

1/ The actual area planted to winter wheat was not determined since significant areas were grazed off when State farms that had been planted to winter wheat were privatized.
2/ Fodder excludes maize and alfalfa.
3/ Measurement bias for mountainous areas is a result of areas being determined from a flat map.

Table 5.7
Albania: 1993 winter wheat area, yield and production

Stratum	Area harvested (ha)	Yield (tonnes/ha)	Production (tonnes)
1	76 800	3.6	276 480
2	38 200	3.5	133 700
3	36 600	3.0	109 800
4A	3 900	2.3	8 970
4B	0	0	0
Total	**155 500**	**3.4**	**528 950**

CHAPTER 6

ARGENTINA

GENERAL PURPOSE AGRICULTURAL SURVEY PROGRAMME BASED ON AREA FRAME SAMPLING METHODS (1972-1997)

1. INTRODUCTION

This chapter describes the general purpose, annual agricultural survey based on area sampling methods conducted in the Province of Buenos Aires since 1993 by Argentina's *Secretary of Agriculture, Fisheries, Livestock and Food* (SAGPyA). This annual survey in the Province of Buenos Aires constitutes a part of the Annual National Agricultural Survey Programme (ENA) conducted by the National Institute of Statistics and Censuses (INDEC). In all other provinces of Argentina, under the responsibility of INDEC, the survey designs are based on list sampling methods using INDEC's 1988 National Agricultural Census as the list sampling frame.

From 1972 until 1986, the Secretary of Agriculture, Fisheries, Livestock and Food conducted annual agricultural surveys in five provinces: Buenos Aires, Cordoba, Entre Rios, La Pampa and Santa Fe. In each of these provinces, a two-stage self-weighted sampling design was used. The PSUs were selected with probability proportional to size and in each selected PSU two segments were selected.

During 1987-1992 the sampling designs were modified in order to improve the precision of the estimates. The new designs incorporated two main technical features: the stratification of the frame into four strata defined by the predominance of agricultural *vis-à-vis* livestock characteristics; and the utilization of one-stage replicated designs. However, during the period 1987-1992 only one survey was conducted.

2. SURVEY DESIGN FOR THE ANNUAL AGRICULTURAL SURVEY CONDUCTED IN THE PROVINCE OF BUENOS AIRES (1993-1997)

2.1 Survey Objectives

As part of the Annual National Agricultural Survey Programme, the objectives are to estimate planted and harvested crop areas; crop production; livestock numbers, by age and sex, at 30 June; acquisitions, sales and movement of livestock; and a number of economic and social characteristics of the agricultural sector.

2.2 The Area Frame Survey Design

The Province of Buenos Aires is divided into *partidos*, which are the smallest political subdivisions. For survey purposes, eight domains were formed by groups of contiguous *partidos* having similar agricultural characteristics. The domains are the smallest areas for which survey estimates and their sampling errors are available.

For the determination of the domains, the density by hectare for each relevant item in each *partido* was calculated and mapped. For example, for cattle, the density per hectare by *partido* was mapped so that similar contiguous *partidos* were formed. This procedure was done for each variable and combined in order to obtain the domains. Finally, eight domains were formed.

The survey design involves a stratified area frame that has, within each of the 8 domains, the following strata:

- A complete enumeration *special stratum*; and
- Four *strata* defined by predominance of agricultural *vis-à-vis* livestock characteristics.

For the special stratum, the reporting unit is the *agricultural holding*. All holdings totally or partially included in the special stratum are interviewed by enumerators in each survey round.

For the area survey design in all the province excluding the special strata, the reporting unit is the *agricultural holding* and the sampling unit the *segment*. Segments are land areas whose boundaries follow holdings boundaries and are delineated on cadastral maps. During the field data collection, the holders of all holdings totally or partially located within the sample segments are interviewed by enumerators.

Except for the area covered by the special strata, a one-stage, stratified systematic replicated sampling design is used.

2.3 Special Strata

The special strata were originally formed by 71 agricultural holdings having large agricultural areas and/or large numbers of livestock. This area covered by the list of special holdings was indicated on the cadastral maps of each *partido*.

During the annual field data collection, these special holding areas (indicated in the maps) are surveyed independently of the changes in ownership between survey rounds.

The purpose of defining the special strata was to improve the precision of the estimates by including with certainty in the survey important agricultural operations that are few in number but that contribute a large share of the total agricultural production.

The table below shows the number of holdings and their areas in the special strata.

Characteristics of the special strata

Domain	Agricultural Holdings	Area (ha)
I	8	85 252
II	-	-
III	8	44 630
IV	10	133 043
V	10	71 679
VI	13	130 809
VII	14	126 323
VIII	8	194 048
TOTAL	**71**	**785 784**

2.4 Stratification in each Domain

Cadastral maps, for each *partido*, in scales of approximately 1:100 000 were used for the area frame construction. The cadastral maps include the delineation of the holdings and the names of the holders, as well as their areas.

Strata were defined in each domain. For this purpose, maps were prepared using satellite imagery from a UNDP project on land use.

A stratum in a domain is formed by large areas with relatively similar agricultural characteristics.

The strata boundaries were delineated (transferred) on the cadastral maps. As already mentioned, the special strata were also delineated on the maps.

Definition of Strata

A: Land predominantly used for agriculture, with some livestock.
AG: Land predominantly used for agriculture, with livestock.
GA: Land predominantly used for livestock, with some agriculture.
G: Land predominantly used for livestock.

2.5 Delineation of Sampling Units (Segments) and Target Segment Size per Stratum

In the cadastral maps, each stratum within each domain was subdivided into *segments*. The segment boundaries follow holding boundaries which are indicated on the cadastral maps.

Segment size (area) varies according to strata. The objective was to define a target segment size that would correspond to an average of three interviews (tracts) per segment.

To determine such a target segment size in each stratum, squares of different sizes were drawn in different points of the stratum on cadastral maps. Then for each square the number of holdings partially or totally included in the square was calculated. A regression study between area and number of holdings determined the target segment size for the stratum. For example, the target segment size for the Stratum GA of Domain V is 400 ha.

2.6 Total Sample Size for the Province

The previous area sample designs provided the basis for determination of the total sample size in the Province of Buenos Aires.

The sample segments of the previous design were delineated on the cadastral maps showing the new strata boundaries. Then, with data for the old segments classified in the new strata, the variance between segments was calculated for the main survey variables. Previously, the data were adjusted to the new target segment size in each stratum.

Then, the coefficients of variation (CV) for the estimator of the total for the main survey variables were calculated, considering also different sample sizes. Analysing such results led to the determination of the total sample size of N = 1 720 segments. For this sample size, it was expected to obtain the following CVs: cattle, 1 %; planted area of wheat, 2.61 %; sorghum, 12.26 %; maize, 4.97 %; soybeans, 2.83 %; and sunflower, 4.24 %.

2.7 Sample Size for each Domain

Once the total sample size was determined, an optimum Neyman allocation determined the sample size in each domain for each of the main survey variables. Then a compromise, considering the different allocations, determined the sample size in each domain as follows:

Sample size

Domain I	120 segments
Domain II	64 segments
Domain III	192 segments
Domain IV	320 segments
Domain V	256 segments
Domain VI	204 segments
Domain VII	360 segments
Domain VIII	204 segments
Total	**1 720 segments**

2.8 Sample Size in each Stratum of a Domain

Following the same procedure as the one used for the allocation of the sample to domains, the sample was allocated to the strata in each domain. For example:

Stratum Sample Size for Domain V

A (land predominantly used for agriculture, with some livestock)	16 segments
AG (land predominantly used for agriculture, with livestock)	72 segments
GA (land predominantly used for livestock, with some agriculture)	112 segments
G (land predominantly used for livestock)	56 segments
Total Domain V	256 segments

2.9 Area Frame Construction

The area frame construction in each domain followed the following steps:

Step 1: Delineation on the cadastral maps of the special stratum and non-agricultural areas. These two types of areas are eliminated from the area frame. In the remaining land, which is the survey area, the strata are delineated in each *partido*.

Step 2: Delineation, on the cadastral maps, of all segments of each stratum, following the determined target segment size. For example, for the Stratum GA of Domain V (target segment size = 400 ha), segments of approximately 400 ha were delineated, trying, whenever possible, not to cut across holding boundaries. As already mentioned, the segment boundaries follow holding boundaries which are indicated on the cadastral maps.

Step 3: The segments were numbered in a sequence that followed the zones of each *partido*, in a serpentine order within each zone. Then, the numbering continued in the next *partido*, until completing the domain. For each stratum there is a sequential numbering of segments through all zones and *partidos* of the domain.

2.10 Sample Selection

The systematic sample selection was done independently for each stratum, and it includes eight replicates. The sample procedure is indicated below:

Assume that the N_{dh} segments in a given stratum h of the domain d have been ordered as described. Then the sample in each replicate should have n_{dh} segments. For selecting one replicate, a sampling interval equal to $N_{dh}/n_{dh} = k_{dh}$ is used. Selecting a random start between 1 and k_{dh} identifies the first selected segment of the replicate. Then adding the sampling interval k_{dh}, all n_{dh} sample segments were selected.

The same procedure was followed for selecting the other replicates, choosing independent random starts.

The use of a replicated design has several advantages: i) provides the possibility of using, for a certain survey round, only some of the replicates, if necessary; ii) allows for sample rotation, which is quite important to avoid interviewing the same holders year after year; iii) unbiased variance estimates are obtained when using two or more replicates.

2.11 Estimation Methods and Calculation of Sampling Errors

Direct expansion weighted segment estimators are those more commonly used. Estimators based on replicates are also utilized. For some variables, a ratio estimator is used for strata with an adequate correlation with the auxiliary variable, and direct expansion weighted segment estimators for the remaining strata.

Estimates and their CVs are calculated by domain and for the total province. The formulas used for the direct expansion area sample estimators and their variances are presented in Chapter 11.

3. DATA COLLECTION

The annual field data collection is conducted as part of the Annual National Agricultural Survey Programme (ENA).

The segment enumeration does not involve objective measurements of areas. However, sketches are used to delineate the area of each holding included in the sample segments.

3.1 Selection and Training of Field Personnel for the 1996 Survey

Field personnel are selected from professionals or advanced students of agricultural or veterinary sciences, with a good knowledge of an area of the province and preferably having experience in agricultural censuses or survey data collection.

After a first screening of candidates, a further selection through personal interviews identifies those to attend the training course. Supervisors and enumerators are selected by evaluating their performance in the course and at the final examination.

Currently, a number of field personnel have been employed regularly for the annual survey rounds. However, all field personnel work is evaluated after each survey round. Those enumerators who do not meet the standards are eliminated, and those performing exceptionally well are promoted to supervisors.

Separate training courses were organized for "new" enumerators and for enumerators with experience of at least one survey round. And, similarly, for supervisors. Experienced supervisors took an active role in the courses.

The training courses emphasize interviewing techniques and the need to cover the sample segments completely.

3.2 Field Organization for the 1996 Survey

The field staff consisted of a total of 12 supervisors (each working in a zone) and 46 enumerators - an average of 3.8 enumerators per supervisor. In addition, there were two subcoordinators from the Provincial Statistical Office and two from the Secretary of Food, Agriculture and Livestock, responsible for training and control of field operations.

Each enumerator conducted an average of 40 farmer interviews. This average was lower for "new" enumerators.

Around 80% of the field data collection was finalized in 60 days. When farmers were not immediately available or residing elsewhere, more time was allowed to finalize the interviews.

The field data collection was initiated the day after the completion of the training course (the first days of July 1996).

However, the field data collection was actually undertaken from the beginning of July 1996 until September, overlapping an important period of agricultural work. This proved to be a problem since, for example, livestock had to be recorded for 30 June 1996, too far from the date of data collection.

The data collection operation involved:

Supervision. Supervision in each zone was undertaken by the subcoordinators during the first week. The purpose is for early detection of enumerator errors. From the second week, enumerators are supervised only by supervisors, who are obliged to meet their enumerators at least once a week. Each supervisor has the responsibility of revising and/or correcting the questionnaires and delivering them to the subcoordinator on a weekly basis. While in the field, the subcoordinator also reviews the questionnaires, and introduces the necessary corrections (if necessary by re-interviews) to obtain the final data for each completed segment. The questionnaires are then sent to the Provincial Statistical Office (in the city of La Plata) for additional consistency checks. Finally, the questionnaires are sent to the SAGPyA for data entry and further analyses and consistency checks.

Reminders C. The Reminders C are used to record that it has not been possible to complete an interview in a specific holding. The remainder contains the address where the holder/respondent has to be visited. In cases where the holder is not located by the enumerator in the field, the interview is naturally delayed, thus the reason for identifying such holdings. Supervisors, authorized by the subcoordinators, send by fax the Reminders C to the zones and the Provincial Statistical Office, so that segments with one or more Reminders C are readily identified.

4. DATA PROCESSING AND ANALYSIS

The survey data processing was implemented by the Programme for Modernization of Agricultural Services - Integrated System of Agricultural and Fisheries Information (PROMSA-SIIAP). The stages of the work are described below.

4.1 Reception and Control of Survey Questionnaires, by Segment, during Data Collection - Data Entry, Consistency Tests and Correction Procedures

Data entry and consistency used the software FOXPRO. Listings of errors were obtained and, after corrections, an updated database was created.

4.2 Data Analysis, Calculation of Estimates, Standard Errors, Correlations and Covariances

The SAS software was used for the above purposes. The statistical analysis, by survey variable, included the following topics: Land tenure, distribution and land use; Cattle and other livestock; and Commercialization.

4.3 Data Processing

Data processing was undertaken using a Micro VAX 3100-80.

The following were obtained before publishing the data:

- Summaries of segments and direct expansion estimates by stratum, domain and for the total province.
- The ratio estimates by stratum, domain and for the total province.
- The frequency of each survey variable.

5. RESULTS FROM THE 1996 ANNUAL AGRICULTURAL SURVEY

The survey estimates are published by domain and for the total province. The table below shows some of the results obtained using direct expansion estimators.

Province of Buenos Aires
Selected Results from the 1996 Annual Agricultural Survey

Item	Estimate	Unit	CV
Cattle (on 30/6/96)	18 221 662	head	2.27%
Sheep (on 30/6/96)	2 231 267	head	8.55%
Wheat planted area	3 620 226	hectares	3.80%
Maize planted area	1 351 194	hectares	4.90%
Cereals planted area	5 395 756	hectares	3.10%
Sunflower planted area	2 087 359	hectares	4.95%
Soybeans planted area	1 449 296	hectares	5.84%
Oilseed crops planted area	3 564 952	hectares	3.69%

6. RECOMMENDATIONS FOR IMPROVING THE SURVEY DESIGN

There are a number of important improvements that can be introduced in the overall survey design. For instance: i) *stratification of the area frame* - appropriate use of high resolution satellite imagery will improve the ability to separate areas with different proportions of cultivated land; ii) *definition of the segments* - by using recognizable physical boundaries; iii) for *data collection* - the use of large-scale aerial photography in place of cadastral maps will make it possible to measure land in the segment, thereby taking full advantage of area sample procedures; iv) a *list frame* component of important holdings should be developed from the next National Agricultural Census together with a procedure for updating the list; and v) the *dates and frequency of the survey data collection* should be adjusted so that timely estimates for winter and summer crops are obtained.

Figure 6.1

Province of Buenos Aires: sample segments

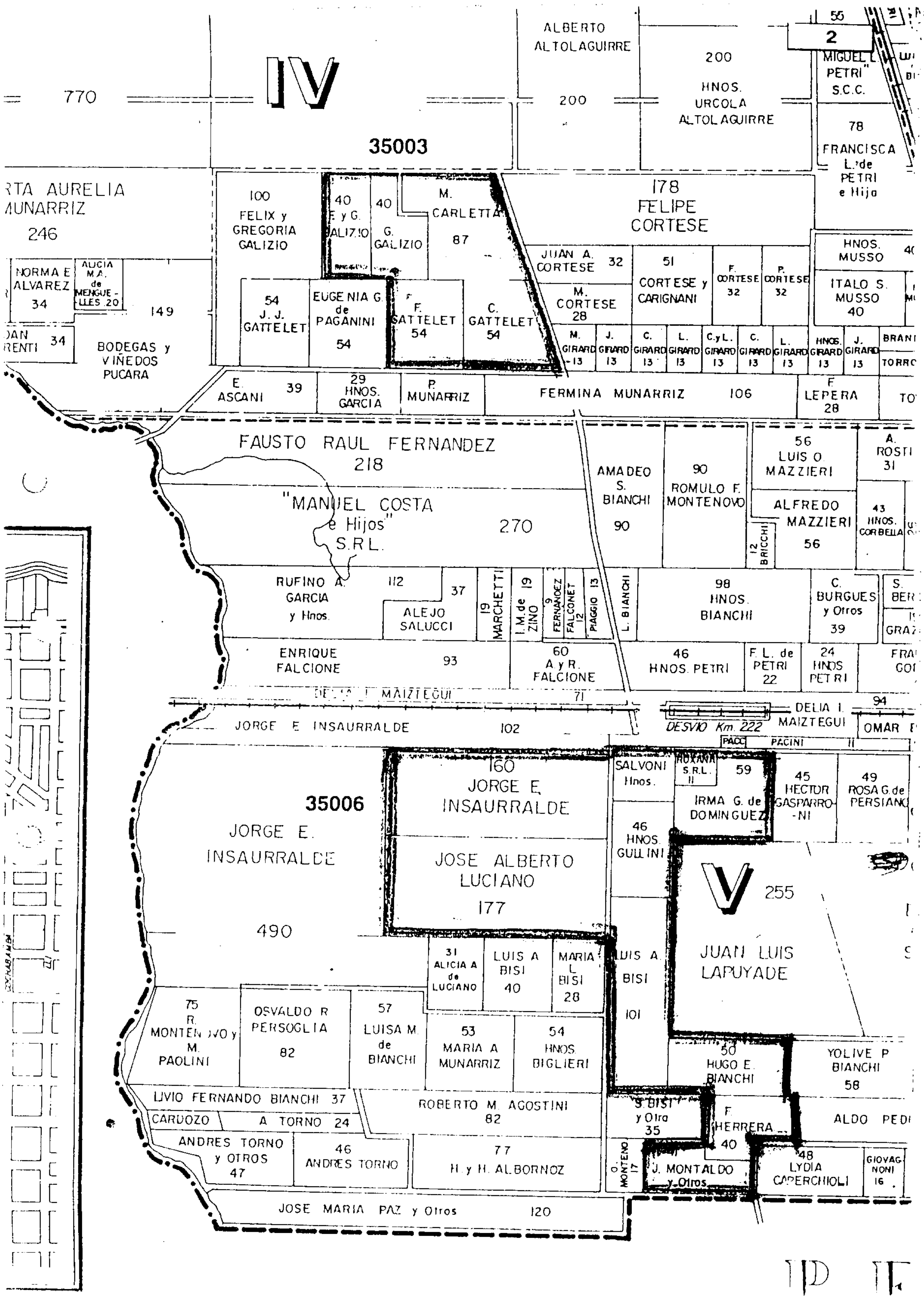
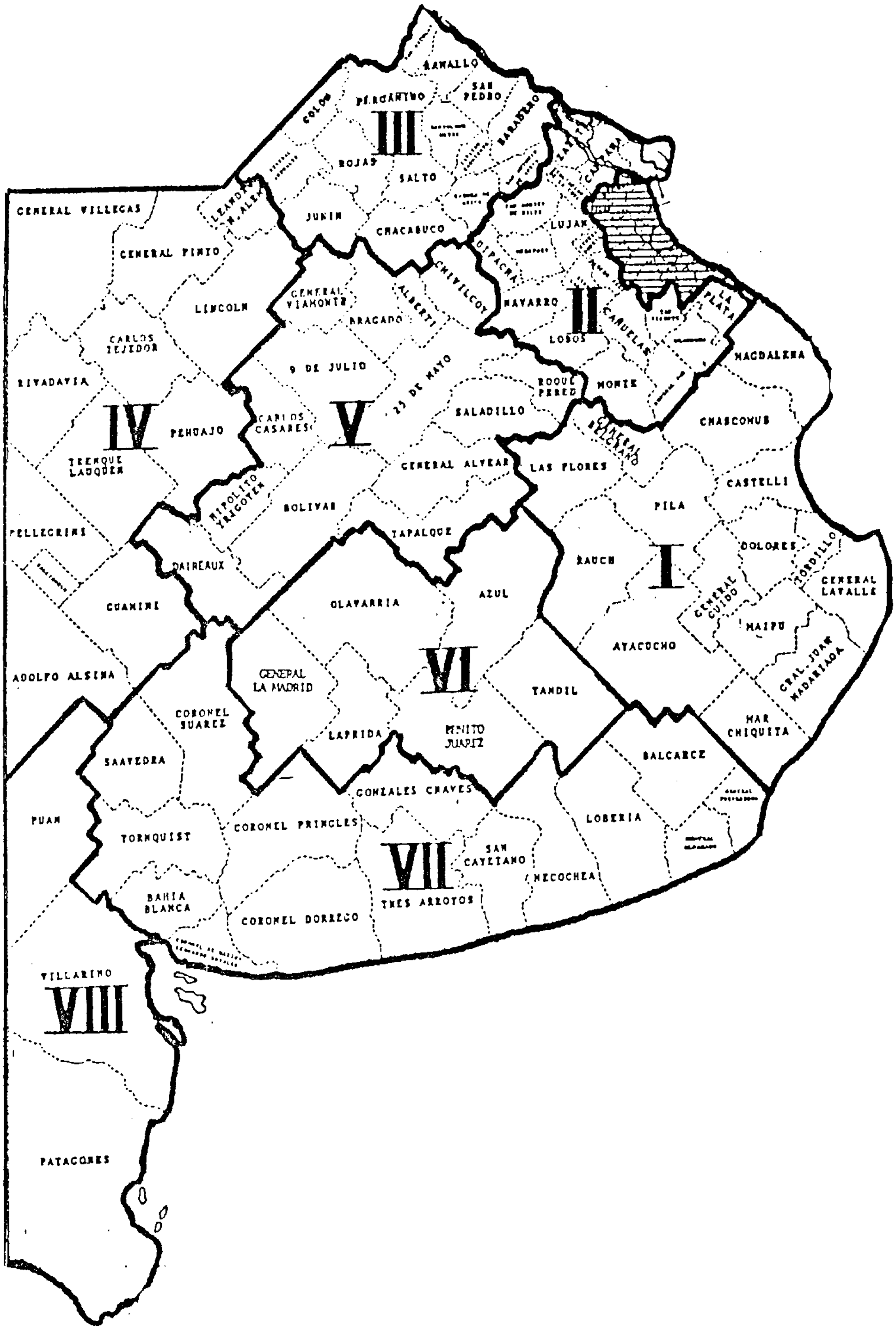


Figure 6.2

Province of Buenos Aires: survey domains



CHAPTER 7

MOROCCO

GENERAL PURPOSE AGRICULTURAL SURVEY PROGRAMME BASED ON AREA FRAME SAMPLING METHODS (1981-1997)

1. INTRODUCTION

This summary describes the methods and practices of the Agricultural Survey Programme based on Area Sampling Methods conducted by the Division of Statistics and Computer Science of the Directorate of Planning and Economic Affairs, Ministry of Agriculture and Agricultural Development (MAMVA). The sector data collection system concerns data for the elaboration and analysis of agricultural policy as well as the information necessary for the efficient operation of a market-oriented agricultural economy.

2. CHARACTERISTICS OF THE AGRICULTURAL SECTOR

The total population of Morocco was around 26.7 million inhabitants in 1966, 50% of whom reside in the rural areas. More than 5.5 million people are employed directly in agriculture or in agricultural support activities. About 9.2 million ha of Morocco's total area of 71 million ha is *usable agricultural land* (SAU).

The main crops are: cereals, occupying about 5 million ha (55% of the arable land); legumes (beans, peas, lentils, etc.), using about 8% of the SAU; olive groves occupy some 418 000 ha.; vegetables, 235 000 ha.; sunflower 110 000 ha.; and citrus 74 000 ha. Livestock breeding is also important. There are more than 3 million head of cattle, 13 to 16 million head of sheep and around 5 million head of goats in Morocco. Horses, donkeys and camels make up about another 1.7 million head. Poultry meat production has grown rapidly from 36 000 tonnes in 1970 to 180 000 tonnes in 1995 and egg production has also increased.

The agricultural sector plays an important role in obtaining foreign exchange by exporting citrus, vegetables, oil from olives and oilseed crops, beans, peas, lentils, etc. Agriculture provides from 18 to 20% of the Gross National Product. The application of irrigation, improved crop varieties and modern crop technology have brought about a significant increase in agricultural production; however, a large portion of the SAU still suffers from periods of drought.

3. DEVELOPMENT OF SURVEY METHODS

3.1 History of Agricultural Statistics in Morocco

There have been four principal steps in the development of agricultural statistics in Morocco:

Before 1974. The collection of agricultural data was carried out by the Secretariat of Planning. Very few surveys were done due to the small amount of funds allocated to this activity.

1974-1977. In 1974, the production of agricultural statistics was transferred to MAMVA with the responsibility of building a permanent data collection system at the regional level. A number of regional bureaux were created. In this period, various sampling designs were employed: a) a sample inherited from the Secretariat of Planning, based on the 1962 Census and a list of those paying an agricultural tax; and, b) perceiving the weakness of the previous frame and sample, in 1976-77 the MAMVA mounted a huge effort to build a complete list of agricultural operations by villages. This produced a large data bank that contained data useful for stratification and selection of a sample of villages.

1978-1983. Beginning in 1978, the new sample - a list sample of villages - from the data bank created in 1976-77 became operational for all agricultural surveys. However, that frame and sample did not serve for as long as was expected. The problems were difficulty in updating and rotating the sample and confusion and instability surrounding the definition of a *douar* (village) used as the sample unit.

1983-1997. Over the years, it became evident that agricultural development was hindered by scarcity of data and doubts concerning the quality of the data. To fill these gaps, the MAMVA requested assistance from the United States Agency for International Development (USAID) and the United States Department of Agriculture (USDA) to implement a new survey system based on area sampling methods.

3.2 Methodology applied after 1983

The surveys conducted after 1983 used two different sampling frames:

i) For all agricultural variables related to land use and crop production, an *area sample* has been used. The area sampling frame currently covers more than 90% of the country and practically all land with a high potential for agriculture.

ii) For agricultural surveys dealing with social and economic development, economic studies and livestock, the *douar list sample* has been the primary frame used. In some cases, the area sample was also used instead of the *douar* list sample: the open segment estimates provide information for an entire farm and can also provide estimates for socio-economic studies, livestock, etc. It is expected that the area sample will replace the *douar* list sample when it is completed for all parts of the country.

3.2.1 List Sample Surveys

The list of *douars* is on computer tape. The record for each *douar* carries 24 characteristics, such as: amount of usable agricultural land, planted areas and number of livestock.

The frame is stratified according to agricultural land use. Sample size was calculated based on usable agricultural land and distributed optimally among the strata. The distribution of the sample among subregions was proportional to usable agricultural land. Sampled *douars* were then selected with equal probability.

3.2.2 Area Sample Surveys

The application of the area sampling methods began as a trial basis in Kénitra Province in 1980 and continued, progressively, until 1996 when all provinces where agriculture is important were covered by an area frame.

The survey design is basically the same as the main survey design described in Volume I of this publication. The following are particularities of the survey design for Morocco:

- A livestock grower is any farmer having at least one head of cattle, sheep, goats, camels or horses.
- A province is the principal political subdivision of the country and forms the level of estimates for agricultural statistics.
- Within a province, strata were divided into two subregions: one containing the irrigated land and the other containing the land without irrigation. The Provincial Directorate of Agriculture (DPA) works in the subregion without irrigation and the Regional Office of Agricultural Development (ORMVA) has the responsibility for the subregion with irrigation.

Stratification. Stratification of the area frame was carried out on a variety of cartographic materials depending on what was available for an area when the work was being done. The stratification for the provinces of Kénitra, Khémisset, Fès, Rabat and Casablanca was done on MSS/Landsat satellite images and maps. For other areas, the stratification was done by using aerial photography and maps of various ages and, in some instances, the frames were stratified using only maps. This required a significant amount of fieldwork to verify the stratification done with old photography and maps. Final stratification was done on 1:50 000 topographic maps or, in a few cases, on 1:100 000 maps. New photography was taken in 1994 and 1995 for the Agricultural Census. New frames and samples were prepared using new photos for each of the regions where old photos had been used. At present, all the stratification is based on recent photography and/or recent TM and SPOT images. For constructing the strata, it was decided that no land use type within a province could be considered a stratum unless it represented at least 5% of the total area. Table 7.1 includes the stratum definitions.

Counting units. For the area frame construction, the next step was to delineate the *counting units* (CU). The CUs were digitized with specialized software (Multi-map, Cries, Erdas) to determine their area and, finally, the entire area covered by the frame was digitized. The software also assigned segments to each CU and ensured that all the frame was digitized. Currently, the frame contains 45 000 CUs with a total area of 530 000 km^2.

Sample selection. A replicated sample is selected from each stratum at the province level. Calculation of sample size and its allocation to strata was done by the software Allocate, developed by the National Agricultural Statistical Service of USDA. The national sample consists of 3 500 segments. Approximately 75% of the sample is allocated to Strata 10, 20, and 30; 15% to Strata 40 and 50; and 10% to Strata 60, 70, and 80.

For the Province of Taounate, detailed frame and sample data are shown in Table 7.2.

A replicated random sample of CUs with probability proportional to size was selected, as shown in the following example.

Assume that the sample is to be selected from Stratum 10 in the Province of Khouribga. In that stratum, the total number of segments, as determined by measurement of CUs and assignment of segments, is N = 4 879. The size of the sample is set at n = 30. The number of replicates is set at 5, therefore each replicate will contain 6 segments. The selection is done on the cumulative column of assigned segments on the listing of CUs. The figure 4 879 is the last entry in this cumulative column. A sampling interval (I) is calculated by dividing 4 879 by 6 = 813.16 thereby creating 6 substrata, each containing 813.16 segments. All random numbers used in the selection of the total sample, n = 30, will be within the range of the sampling interval (i.e. between 1 and 813.16).

The first CU to contain a segment in the first replicate is chosen by a random number: for example, 665. The chosen CU contains segment number 665 according to the cumulative column. The next CU in the first replicate is chosen by finding *another random number* and adding the sampling interval I to the random number in order to make the selection from the second substratum. If the random number selected was, for example, 413, the CU containing segment number 1 226.16 (413+813.16 = 1 226.16) is chosen from the second substratum.

Another random number plus twice the sampling interval I (1 626.32) arrives at the third segment and so on, until a random number plus 5xI indicates the sixth and last CU to provide a segment for the first replicate. Another random number start is used for the second replicate whose selection proceeds in the manner described for the first replicate. It continues in this way until the 5 replicates are selected. This procedure uses the same type of CU listing sheet described in Volume I but with the described approach to the actual sample selection of one CU per substratum in each replicate.

The indicated CUs are then divided into the assigned number of segments. The segments in the CU are numbered and one is chosen at random with equal probability of selection. It is possible that a CU could be selected more than once for different replicates. A segment will be selected from a CU each time it is chosen in the sampling process.

Selected segments are located and outlined on the topographic maps and measured with a planimeter or digitized. A photographic enlargement, usually in 1:5 000 scale, is ordered to be used to guide the field enumeration. Before a survey, the enumerators go to the field with local authorities to identify segments and fix a date for meeting with all farmers who operate land within the segment.

Estimation methods. Closed segment and open segment estimators are used. For crops and land use data, the closed segment has the advantage of being verifiable using the photographic enlargement. With the open segment, farmers' statements about what lies outside the segment are beyond the control of the interviewer.

Two methods are used to calculate estimates of totals, averages and their variances. One employs replicate totals and the other performs the calculations by substrata. The formulae used are shown in Chapter 11.

4. SURVEY METHODS AND CHARACTERISTICS

4.1 Current Surveys

4.1.1 Land Use Survey

The land use surveys are conducted annually to estimate area planted by crop. The survey is carried out in two phases. The first round takes place from 15 February to 15 April. It gathers data on autumn crops planting and on planting intentions for the spring. The second round takes place in May and June, gathering final planting data for the spring crops.

The sample includes about 70 000 farmers and provides estimates at the provincial level and at the level of "special action zones". The area frame sample covers 90% of the area surveyed, the remaining 10% being covered by the village sample.

4.1.2 Crop Forecast Survey

The crop forecast survey is carried out annually during the latter part of April to obtain estimates of expected yield for the four principal cereal crops: hard wheat, soft wheat, barley and maize. It employs a subsample of the area sample used to gather the data to estimate area planted and is done immediately after the planting survey. This survey is of great interest because it provides information about the expected harvest of the current crop cycle.

4.1.3 Crop Yield Surveys

The land use survey provides a list of fields that is used to select a self-weighting subsample of fields of cereals and legumes for the crop yield surveys. The crop yield surveys are carried out from May to September, with the exact date depending on the time of maturity of each crop. Samples of the mature crops are harvested and sent to a laboratory for threshing, assessing moisture content, etc. to make an objective estimate of yield. The estimated yields are combined with the areas provided by the land use survey to determine production by crop. The final results are published by province and by subregion. As with the land use survey, most of the fields sampled are from the area sample; less than 10% are from the village sample.

4.1.4 Olive Survey

The main objective of this survey is to estimate olive and oil production. It is done in a manner similar to the yield surveys for other crops and takes place only in the olive-producing areas. An area frame sample is used.

4.1.5 Price Surveys and Livestock Surveys

Three types of surveys relative to the price of agricultural products are conducted each year by the Division of Statistics and Information.

- A "prices paid to producers" survey takes place each year in October-November to obtain prices paid to producers for each product sold. This survey is done at the same time as the livestock survey, using the same sample. The sample is selected in approximately equal proportions from the two sampling frames used for surveys in the country.
- A "market price" survey is done in selected markets to follow the changes in prices of the principal cereals and legumes. A "wholesale market" survey gathers price information from the main wholesale markets for the important fruits and vegetables.
- Livestock surveys are conducted twice each year, in March-April and October-November. Their main objective is to obtain an inventory of each herd by age and sex with additional information on births, deaths, slaughter, sales and purchases. The sample includes around 40 000 livestock growers.

4.2 Special Surveys

In addition to the regular (current) surveys, a number of special surveys have usually been conducted to meet special requests from the government. Examples of these surveys are project progress surveys, the census of citrus, the vineyard census, slaughter for special occasions such as Aïd al Adha, and others.

4.3 Agricultural Census

In 1996 a general Agricultural Census was conducted at the national level. The fieldwork required six months. Data entry was to have been completed by December 1997.

4.4 The Area Sample Survey Organization

Once the sample and survey materials are ready, a training session is held before start of the fieldwork. This is a plenary session attended by all the service chiefs and their collaborators and designed to inform them of errors made in the previous survey as well as of changes in the sample and questionnaire, and to give them a general idea of the survey programme for the current year.

The enumerators are organized into teams with two or three enumerators who have the responsibility of gathering the data from farmers. The team is led by a team chief who, besides being an enumerator, is charged with the first review of the data collected. He must review all questionnaires completed by the team.

The chief of service and his assistant supervise the surveys, make unscheduled checks and keep the central office informed about the progress and problems, if any, during the survey.

Control of quality is maintained in several ways. One way is the sending of experienced personnel to work with an enumeration team. This is particularly effective in the early days of a survey when it is possible to detect misunderstandings and errors in time to make corrections. Unannounced check visits to segments already enumerated is employed as another control procedure. In addition to the review of each questionnaire by the team leader, a random sample of 20% of the completed questionnaires from each team is reviewed by a special group to check for errors.

Data entry is performed at both the regional and national levels. Data are entered by one person and re-entered (verified) by another. Data entered at regional offices are sent to the central office for summarization at the national level. The use of modems for data transmission is being considered. A generalized edit and summary system using SAS manages the data and provides the results.

5. CONSIDERATIONS FOR FUTURE SURVEY DESIGNS

This summary provides a general view of the statistical information system of Morocco and should serve as a guide for the development of survey designs for either list or area sample surveys. It demonstrates the coexistence of a list sample and an area sample that often complement each other and sometimes substitute for each other when required by technical and/or budgetary restrictions.

The following are some possible considerations for future survey designs:

- Progressive privatization should be accompanied by an allocation of enough funds to MAMVA in order to ensure a better manner of data collection and dissemination of both basic and analysed data.
- The data collection system should furnish both regional and national data in a timely manner.
- To respond to an increasing demand for more specific information, the collection of statistical information should also include, besides the normal variables of crop area and production and livestock inventories, the necessary economic data needed for analysing the agricultural policies and maintaining contacts with the economic conditions of the sector.
- Regular cost of production surveys and rural households studies should become a part of the national data collection system.
- A permanent system for the prompt publication of survey results and for the management of an agro-economic information data bank should be set up.
- In order to estimate properly the increasingly important specialized crops, the number of segments in the national sample should be increased, special list frames should be developed and multiple frame sampling methods should be applied.
- A general census of agriculture should be conducted to provide data at the level of the smallest administrative units. The census could provide the base for developing a Multiple Frame Agricultural Survey Programme and provide specialized list sampling frames.
- The weighted segment estimator should be used to improve survey results.
- The information handling capacity of regional offices should be improved so that data entry, summarization and publication can take place more rapidly.
- Further efforts should be made to become acquainted with statistical agencies in other countries; new procedures should be adapted to the conditions in Morocco; and other ways of improving survey methodology should be established.

Table 7.1

Morocco
Stratum definitions with counting unit and segment sizes

Stratum	Stratum Description	Counting Unit Target Size and Range (ha)	Segment Target Size and Range (ha)
10	Cropland without irrigation	750 ± 10	50 ± 5
20	Irrigated cropland	750 ± 10	50 ± 5
30	Orchards - fruit, nuts, olives	750 .± 10	50 ± 5
40	Forests	2 000 ± 10	200 ± 10
50	Grazing and other non-cultivated land	2 000 ± 10	200 ± 10
60	Small towns	40 ± 5	2 ± 0.25
70	Large towns	20 ± 5	1 ± 0.25
80	Large villages	60 ± 5	4 ± 0.50

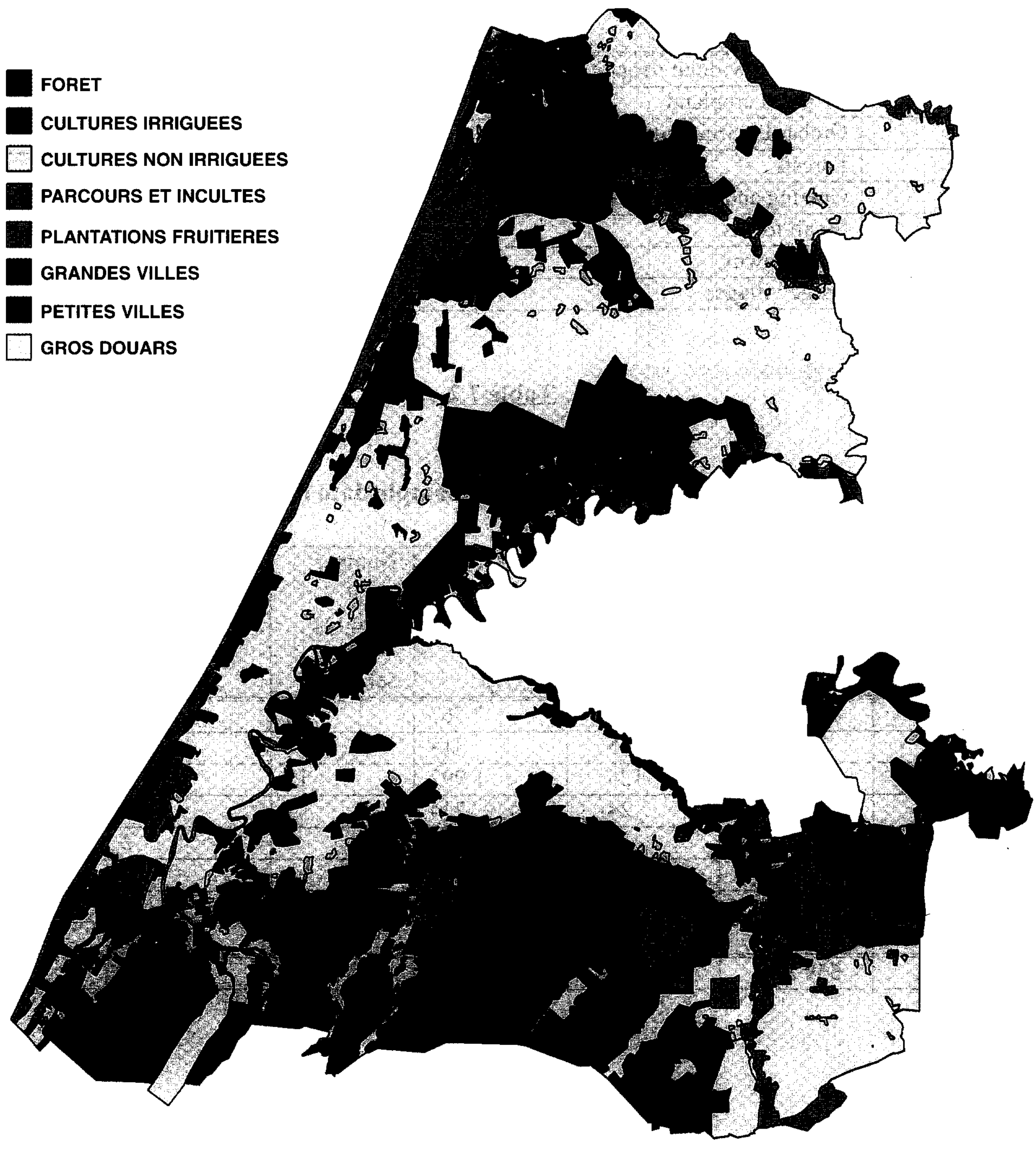
Table 7.2

Morocco
Taounate Province: frame and sample data by stratum

Stratum	Total area (ha)	Total number of counting units	Total number of segments (N)	Number of counting units selected	Number of replicates (R)	Number of sample segments (n)
10	388 364	452	7 767	40	7	42
20	15 919	25	318	4	2	4
30	99 585	124	1 992	13	4	16
40	46 730	24	234	5	3	6
50	7 399	4	37	3	2	4
60	262	5	131	4	2	4
70	178	20	178	4	2	4
80	916	21	229	3	2	4
TOTAL	**559 353**	**675**	**10 886**	**76**	**24**	**84**

Figure 7.1

Land use strata - Province of Kenitra



CHAPTER 8

NICARAGUA

GENERAL PURPOSE AGRICULTURAL SURVEY PROGRAMME BASED ON AREA FRAME SAMPLING OF AGRICULTURAL HOLDINGS USING POINT SAMPLING METHODS (1994-1997)

1. INTRODUCTION

Nicaragua is a nation in transition, struggling to manage large economic and social change with virtually no guidance from historical precedent. The agricultural sector suffered for many years from lack of information about agricultural inputs and production. The Ministry of Agriculture (MAG) tried with limited resources and lack of technical experience to collect area and production data using non-probability surveys. The Central Bank, which is responsible for the National Accounts, found the MAG data unreliable and began generating its own estimates. The lack of agreement between these two series of estimates created confusion and anxiety for members of the Economic Cabinet. Because of incomplete and poor statistics, the Cabinet could not reach a consensus on a plan for the agricultural sector. This led to a request to the National Agricultural Statistics Service (NASS) of the United States Department of Agriculture (USDA) to assist in establishing a procedure for providing a series of "official" estimates. A periodic Area Frame Agricultural Survey Programme was established and surveys have been conducted regularly since 1994.

2. SURVEY DESIGN AND FRAME CONSTRUCTION

2.1 Survey Objectives

The objective of the survey programme was to provide estimates of area and production for basic grains in seven regions and 16 departments in Nicaragua. The basic grains are maize, dried edible beans, sorghum and dryland rice. Cattle data were also collected. The Atlantic region was excluded from the study area because of its relatively minor impact on the agricultural sector.

2.2 Selecting a Sampling Design

Nicaragua had been using a list frame for its non-probability surveys whose quality and completeness were not known. Developing a quality large-scale list frame would have required more time, resources and technical capability than were then available in Nicaragua. The alternative was to develop an area frame that would provide complete coverage of the population to be sampled. An area frame is durable and can be used to gather all types of agricultural information associated with a farm or the land.

2.3 General Characteristics of the Area Survey Design

Topographic maps in a scale of 1:50 000 were available for the entire study area to form the base for the construction of an area sampling frame. No funding was available for the purchase of satellite images or photography necessary to identify boundaries, so an area frame with strata, counting units and segments with identifiable physical boundaries could not be constructed. It was therefore decided to use a stratified, two-stage, area point sample design with probability proportional to size (PPS), and a *segment* (last stage sampling unit) to be *the land area of a holding*. The data collection procedure in each selected holding does not include objective measurements of areas, thus enumeration costs are reduced. The segment boundaries are identified at the time of the interview. The holding (farm) in which the sample point is located becomes the segment or sample unit. During the field data collection, the enumerator must account for all the land inside the segment (farm) boundaries.

The goal was to generate estimates with a margin of error of less than or equal to 10% at a confidence level of 95%. The advantages of this type of point sampling design are: the time required for development and the cost of materials are considerably lower than those needed for an area frame whose elements have identifiable physical boundaries; and less maintenance is required of the topographical base maps.

2.4 Stratification

The stratification consisted of the delineation on maps of land areas into land use categories. Whenever possible, stratum boundaries were placed following identifiable physical features. The stratification was done by cartographers at the MAG on 1:50 000 scale topographic maps. Ministry personnel with field experience assisted in identifying the strata. Satellite imagery would have greatly improved the stratification but, unfortunately, it was not used because of the cost. Table 8.1 shows strata definitions and areas by stratum for the seven-region survey area.

2.5 Primary Sampling Units (PSUs)

In order to ensure proper geographic distribution across the department, each land use stratum was subdivided into uniform *blocks* of a size convenient to be measured with a planimeter. The blocks were ordered in a serpentine fashion beginning in the northeastern corner and proceeding from east to west. Cities, other towns and non-agricultural land (Strata IV, V and VI) were not sampled. Each block was numbered and the area within the block measured with a planimeter.

Almost all producers within a department's stratum are very similar in their production characteristics. This means that the segment size can be fairly small and still be representative of the population of interest.

2.6 Segments. Sample Size and Allocation

The *segments* (second stage sampling units), as already mentioned, were the land area of the holdings. Taking into account the required precision of the estimates, data from previous non-probability and panel surveys were used to calculate the initial sample size and allocation. The sample allocation was to strata within departments. The initial sample size for the seven regions, first used in the 1994 survey, was 5 600 points (holdings). The same sample has been used since 1995. See Table 8.2 for the sample allocation by stratum, department and region.

3. SAMPLE SELECTION AND ESTIMATION

The time required for frame construction and sample selection for the seven regions with a total area of 70 535 km^2 was approximately three months. Planimeters were used for measuring land areas. Sample selection was done manually using grids and random number tables.

In the first stage of the stratified two-stage sampling design adopted, blocks were arrayed and measured as described above and listed. An accumulative column of area (in km^2) was used to select a systematic sample of blocks. The total area for a stratum within a department was divided by the number of sample points allocated to calculate a sampling interval. From a random start, the systematic application of the sampling interval to the column of accumulated area indicates the blocks in which random points were to be located. The number of points assigned was proportional to the size of the block - practically the same procedure as the selection of counting units to contain segments in the traditional area frame process.

For the second stage, which is the selection of points (holdings), a grid was used with 51 points along one side and 65–points along the longer axis. The area under the grid contains 3 315 points. One size of the grid measures 195 mm by 163 mm. A larger grid with the same point spacing was prepared for larger blocks. For sampling, the grids are printed on clear transparencies. The grid is positioned over a selected block with its top edge at the northernmost point or edge of the block and situated as far to the east as possible, still covering the block and maintaining a north-south orientation. A table of random numbers is used to select coordinates (x,y) that indicate the sample point. The coordinates were marked on the grid and the random points were marked directly on the topographic map.

Association of points and farms. The procedure used for selecting random points occasionally led to points falling in non-farm areas such as roads, streams, property boundaries, etc. When this occurred, enumerators were instructed to enumerate the first farm found to the north of the selected point. If the point fell on the boundary of two farms, the farm to the northeast became the sample unit. However, a body of non-farm land use, such as a school, a churchyard, a cemetery, a military reserve, was considered a valid segment for which land area and land use must be accounted. A farm can consist of non-contiguous parcels of land as long as the operator considers them as one farm and they meet the common criteria of using the same equipment and labour and under the same management (cf. Volume I).

3.1 Estimation Methods

The probability of selection is equal to the area of the sampled farm divided by the land area represented by each sampled point. Unequal probability of selection was therefore expected.

In area point sampling, all possible coordinates within the stratum represent all possible sampling points.

For each farm (unit) in the population, measures of size, x_i, may be available for each unit "i" in the population, and x_i may be expected to be approximately proportional to a variable, y_i, to be measured in a survey. For example, x_i may be total farm area, and y_i may be production expenses in a given year. The procedure of PPS sampling will then permit efficient estimation of the population total of the y_i. PPS sampling may be viewed as a limiting form of size-stratified sampling. In PPS sampling with replacement, the sample is formed by n independent draws, at each of which the i^{th} unit of the population is selected with probability $p_i = x_i / X$ where $X = x_i + \text{ - - - } + x_a$. In this case, the estimator for the population total, $Y = y_i + \text{ - - - } + y_n$, is:

$$\hat{Y}_{wr} = \sum_{i=1}^{n} y_i / np_i$$

This estimator is unbiased with variance

$$V(\hat{Y}_{wr}) = n^{-1} \sum_{i=1}^{N} p_i (p_i^{-1} y_i - Y)^2$$

An unbiased estimator of $V(\hat{Y}_{WR})$ is

$$\hat{V}(\hat{Y}_{wr}) = [n(n-1)]^{-1} \sum_{i=1}^{n} (p_i^{-1} y_i - \hat{Y}_{wr})^2$$

The following illustration shows how farm-level data are expanded to represent the population of interest. The figures used could correspond to a stratum within a department (specific area). A sample allocation is shown with the area represented by each sample unit if the segment size was equal to one unit of measurement. In this case, the area units are *manzanas* (1 *manzana* = 0.7 ha).

Stratum	Manzanas	Sample Points	Manzanas per Point
I	41 160	84	490

The probability of selecting a 7-*manzana* farm in Stratum I (high cultivation) and its sample weight from the information above is:

Probability of selection = 7 / 490 = 0.0142857
Sample weight = 490 / 7 = 70

For this 7-*manzana* farm, the expansion factor is 70. If the 7-*manzana* farm has 2 manzanas of beans, 4 manzanas of maize, 4 manzanas of pasture, and 3 head of cattle, the expansions would be as follows:

Total Expanded Area = 7 x 70 = 490 manzanas;
Bean Expanded Area = 2 x 70 = 140 manzanas;
Maize Expanded Area = 3 x 70 = 210 manzanas;
Pasture Expanded Area = 2 x 70 = 140 manzanas;
Number of Cattle Expanded = 3 x 70 = 210 head.

The summation of these expanded variables for all 84 sample points provide the estimate of each item for Stratum I.

4. DATA COLLECTION

4.1 Selection and Training of Field Staff

Eighty enumerators are hired for each survey round. The enumerators are part-time workers, generally hired from the same area in which they will enumerate. They receive three weeks of training. The first week is exclusively devoted to classroom instruction on reading topographic maps; the second week consists of field exercises on locating selected points; and the third week of training covers questionnaire content and interviewing techniques.

4.2 Field Data Collection

Enumerators were provided with topographic maps (scale 1:50 000) with the selected points marked to identify the area they are responsible for enumerating. Each enumerator was expected to complete an average of 70 interviews. The first step in enumerating was to locate the point and identify the farm. Next, they must find the farm operator or a knowledgeable informant to complete the questionnaire. Finally, the enumerator makes a sketch of the farm on an appropriate page in the questionnaire. Landmarks, instructions for arriving at the farm, the location of the random point and other pertinent information are also recorded.

The supervisors are trained cartographers. Each supervisor is responsible for four enumerators. They assist in point location and resolve data collection problems. Supervisors are also responsible for quality control checks on approximately 20% of their enumerators' assignments.

The fieldwork for the first survey covering the seven regions with a sample size of 5 600 points was completed in about 30 days.

Since 1996, two surveys are conducted each year, the first in March/April and the second during September.

5. DATA EDITING AND SUMMARIZATION

Field supervisors reviewed all questionnaires prior to sending them to the central office in Managua. Data inconsistencies were reviewed and resolved. Notes were written explaining irregularities. The MAG staff reviewed the questionnaires for legibility and assuring the presence of the required entries.

The data from each questionnaire were entered using KeyEntry III software especially designed for code-data keying. Each data cell had a unique code identifying the item. Each code-data pair was entered into the computer. This facilitated computer editing and summarization. The software also permits data verification.

The data were computer edited using the Lan Xbase Edit System (LXES). LXES is a NASS developed system using CLIPPER software. A sample master computer file containing all the sampled points was generated. Incorporating the sample master into the computer edit assured that all the sampled points were tabulated or otherwise accounted for. Edit checks were programmed into the system to flag data irregularities such as extremely high or low yields. These were classified as warning errors, meaning that computer data were to be compared against the questionnaire. This assured that the warning error was not due to data entry error.

Data irregularities identified as unrealistic, such as harvested area reported with no production or vice versa, were called critical errors and had to be corrected before a record was summarized.

After all data had been computer edited and deemed "clean", further analysis was conducted. Listings of potential outliers and records where expanded data was not similar to others in the same stratum in that department were generated and reviewed. In some cases, questionnaires were returned to the field for verification of questionable data.

Even though the response rate was 96% in the last surveys, the data were adjusted for non-response. The procedure consisted of adjusting the area represented by each point based on the number of usable reports (usable reports are defined as all reports completed by the enumerators).

Once corrections were made, sample weights were calculated for each record as described earlier. The data were then summarized using a Foxpro summary system. Many analytical tools were written into the system such as the number of positive reports, the mean of positive reports, direct expansions and coefficients of variation.

6. SURVEY RESULTS

Estimates were published at the department level (16 departments) for maize and beans. Rice and sorghum were published at the regional level (seven regions). Cattle data were published as a total estimate for the seven regions. Table 8.3 shows selected results (for seven regions) from the 1995 Nicaraguan Agricultural Survey.

7. ACTIVITIES TO DATE

A pre-test of procedures was carried out during the summer of 1993. The ability of enumerators to locate points and manage a proposed questionnaire was tested in selected locations throughout the country. After the pre-test showed the process to be feasible, a frame and sample were prepared for Region 4. A pilot survey to obtain cost of production data in Region 4 was carried out during the latter part of 1993. Both the pilot survey and pre-test results supported the practicality of the proposed area point sampling method. An area and production survey covering the seven regions was carried out using the described point sampling method in September 1994. From 1995 to 1997, similar surveys were conducted in March and September to gather data on basic grains.

8. RECOMMENDATIONS AND CONCLUSIONS

Area sampling frame development is a major undertaking that must be considered a long-term investment. The efficiency of the frame over time depends on the strata definitions and the quality of the stratification. When land use does not change, restratification is not necessary. However, the stratification could be improved with the use of high-resolution satellite imagery. As land use changes and the scope of the project increases, restratification may be necessary. Given the limited resources available, the sampling frame constructed was adequate and performed extremely well for estimating basic grain production.

The Ministry of Agriculture conducted a second basic grains survey in 1995 to measure production for the second and third season crops, thus providing an assessment of basic grain production for an entire year crop cycle. Plans are to continue with the current survey design so that changes in levels of production can be measured. The area point sampling frame will also be used to carry out a cost of production survey for basic grains.

Recommendations call for expanding the survey to include other widely grown crops such as soybeans, groundnuts and melons. A sample rotation design should be incorporated to reduce respondent burden but still allow for a reliable measure of change through the use of the ratio estimator. A multiple frame design incorporating lists of holdings from the planned 1998 National Agricultural Census should be added to the system to better estimate special crops that are highly concentrated and not proportionately associated with land use. A list of large producers of crops and livestock is being prepared for use in future surveys.

Table 8.1

Nicaragua: strata definitions

Stratum	Stratum definition	km²	*Manzanas*1/
I	High cultivation - 60% to 100%	12 400.59	1 771 512.86
II	Medium cultivation - 40% to 59%	27 255.12	3 893 588.57
III	Low cultivation - 1% to 39%	14 069.99	2 009 998.57
IV	Cities	110.57	15 795.71
V	Other towns	259.66	37 094.28
VI	Non-agricultural	16 439.86	2 348 551.42
Total		**70 535.79**	**10 076 541.43**

1/ 1 *manzana* = 0.7 hectare = 1.73 acres.
The non-agricultural stratum includes mountains, lakes and national parks.

Table 8.2

Nicaragua: sample allocation by department

Region	Department	Stratum I	Stratum II	Stratum III	Total
	Nueva Segovia	132	96	84	312
	Madriz	72	120	24	216
	Estelí	84	156	48	288
I		**288**	**372**	**156**	**816**
	Chinandega	288	108	24	420
	León	300	84	60	444
II		**588**	**192**	**84**	**864**
	Managua	288	144	36	468
III		**288**	**144**	**36**	**468**
	Masaya	252	0	0	252
	Granada	216	24	24	264
	Carazo	168	48	24	240
	Rivas	240	60	36	336
IV		**876**	**132**	**84**	**1 092**
	ZC Oeste	192	168	96	456
	Boaco	120	204	24	348
	Chontales	180	216	32	428
V		**492**	**588**	**152**	**1 232**
	Jinotega	84	204	60	348
	Matagalpa	84	108	72	264
VI		**168**	**176**	**132**	**476**
	Río San Juan	144	84	72	300
IX		**144**	**84**	**72**	**300**
TOTAL		**2 844**	**2 040**	**716**	**5 600**

Table 8.3

Nicaragua: selected results from the 1995 Agricultural Survey

Item	Direct Expansion (*manzanas*)	Coefficient of Variation (percentage)
Land in farm	7 913 484	0.7
Permanent cropland	1 255 042	1.9
Annual cropland	389 515	3.7
Pasture	3 848 361	1.0
Rangeland	1 571 233	1.9
Forest	705 899	2.7
Farmsteads	94 612	4.0
Other areas	48 821	8.3
Maize planted	353 916	3.1
Beans planted	86 257	5.6
Dryland rice	37 310	11.7
Sorghum planted	36 857	11.2

CHAPTER 9

PAKISTAN

GENERAL PURPOSE AGRICULTURAL SURVEY PROGRAMME BASED ON AREA FRAME SAMPLING METHODS (1985-1997)

1. INTRODUCTION

This chapter describes the general purpose Agricultural Survey Programme based on area sampling methods implemented in Pakistan. It also includes a brief section on the Objective Yield Surveys that use a subsample of the area frame.

The survey programme started in 1985 and covered seven districts up to 1988. The district area frames were constructed using a photomosaic of 1976 aerial photography.

From 1989, the survey programme was gradually extended and involved the use of SPOT satellite imagery for assisting the construction of the area sampling frame (AFS) in which strata, primary sampling units (PSUs) and segments have recognizable physical boundaries. This is the largest area sampling frame ever constructed using SPOT satellite images.

The area frame construction and sample selection in more than 90% of the country follow those explained in the appropriate sections of Volume I of this book.

In Special Areas, covering less than 10% of the country, a point sampling method was applied to identify sample segments. These Special Areas were some of the sensitive border areas, or areas for which the photography was not available, or the physical features were rare, and it was therefore not possible to subdivide the selected PSUs into segments.

From June 1994, the National Agricultural Statistics Programme based on area sampling methods, with an area sampling frame constructed utilizing SPOT satellite imagery, has been operational for practically the entire country, with the exception of the terraced areas. The sample selection for the terraced areas is currently being undertaken using a line segment approach.

From 1998, multiple frame methods will be applied in order to improve the precision of several important survey estimates.

The survey programme was designed and implemented by the Federal Bureau of Statistics with technical assistance from the National Agricultural Statistics Service (NASS) of the United States Department of Agriculture, as part of a Government of Pakistan/United States Agency for International Development (USAID) project.

2. DEVELOPMENT OF AGRICULTURAL SURVEYS BASED ON AREA SAMPLING DESIGNS

Prior to 1985, Pakistan estimated agricultural production using the Village Master Sample (VMS). The VMS employed sampling procedures based on the Land Revenue System. It was found that the VMS system of collecting agricultural statistics did not fully meet the requirements of planners and policy-makers because crop estimates were not available on a timely basis and reliability of the estimates was not known.

As a consequence, the Government of Pakistan made a request to USAID for technical assistance in improving the existing system of agricultural statistics. A team from the NASS recommended the implementation of a National Agricultural Statistics Programme based on area sampling methods.

The Area Sample Survey Programme would replace the VMS data collection system by:

- providing statistically reliable estimates of area and production for five major crops, i.e. wheat, rice, cotton, maize and sugar cane, on a timely and cost-effective basis;
- providing a frame to be used for estimating the yield of major crops immediately after crop maturity by harvesting sample plots in selected segments of the area frame;
- providing a frame for sampling that would make possible the use of growth models for forecasting yields of major crops before harvest in order to make early production estimates.

Up until 1997/1998, both the VMS and the ASF survey systems would remain in operation.

Table 9.1 shows a comparison of wheat area as estimated by the Village Master Sample with the estimates obtained from the Area Sample Survey.

The main advantages of applying area sampling methods as the statistical foundation of the National Agricultural Statistics Programme are the following:

- survey data are completely independent of all other sources such as revenue data or censuses, and the area sample method assures total coverage of the area to be sampled;
- surveys can be conducted when needed and results can be produced in a timely manner;
- they are cost-effective, statistically sound, and can be used for collecting any type of data related to land or farms.

Phase I (1985-1988).

In 1985, aerial photography in scale 1:30 000 from 1976 and 1:50 000 topographic maps and planimeters were used to stratify and prepare the *mosaics* of contact prints for an area sampling frame in seven districts: four in the Punjab and three in Sindh Province.

Each stratum block was subdivided into PSUs with a target size of 10 segments, and a size within 2 and 15 segments. Planimeters were used to measure the areas. The total official area of a district could not exceed by more than 5% the sum of all PSU size measures in the district frame. Strata, PSUs and segments had identifiable physical boundaries.

The Phase I survey designs had four active replicates. When the sample size for a stratum within a district and the number of replicates were determined, several simple random samples (replicates) of segments were drawn from that stratum which would sum to the predetermined total sample size. In each substratum *four* segments were selected at random without replacement. A replicate is a simple random sample (SRS) of one segment per substratum. Replicated sampling facilitates sample rotation, sample adjustment, and computation of variances for the data collected. Samples were drawn at the district level which was the target population.

Crop area and objective yield surveys have been conducted continuously in the seven districts of Phase I since 1987-1988. It was not possible, however, to get additional or more recent aerial photography.

Phase II (1989-1994).

This phase of the project began in1989 employing SPOT satellite imagery with 20m resolution to stratify and construct an area sampling frame for the entire country. Surveys in the seven districts of Phase I were continued, using their original frame until the new frame was completed in 1994.

The National Agricultural Survey Programme has been operational since June 1994, except for the terraced areas. The basic features of Phase II sample designs are similar to those of Phase I, except for the change in the number of replicates.

Phase III (1994-1997).

The sample selection for the terraced areas is currently being undertaken using a line segment approach.

Unless otherwise stated, the survey design used since 1989 is described in the following sections.

3. GENERAL CHARACTERISTICS OF THE AREA SAMPLE DESIGN

Pakistan is divided into four provinces (Balochistan, North-West Frontier Province, Punjab and Sindh), which are in turn subdivided in a total of 74 districts.

An independent area sample frame was constructed for each province. A replicated area sampling design was used. Strata were defined by proportion of cultivated land and other land use characteristics. Strata, PSUs and segments had identifiable physical boundaries and an approximately equal size in each stratum. Two active replicates were selected. However, in one province the districts were clustered, ordered and then four replicated simple random samples were selected.

The area frame construction and sample selection in more than 90% of the country followed the procedures described in Volume I of this book.

In Special Areas, covering less than 10% of the country, a point sampling method was applied. These Special Areas were some of the sensitive border areas, or areas for which the photography was not available, or the physical features were rare, and was therefore not possible to subdivide the selected PSUs into segments based on what could be seen on the topographic map.

4. LAND USE STRATA DEFINITIONS AND BLOCKS

For both the Phase I and Phase II designs, the following strata were defined:

Stratum 10 = Intensive agriculture, includes land areas which are 60% to 100% cultivated at any time during the year. Minimum block size = 1 km^2. Delineation and shading is in *purple*.

Stratum 20 = Extensive agriculture, includes land areas which are 30% to 59% cultivated at any time during the year. Minimum block size = 2 km^2. Delineation and shading is in *sky blue*.

Stratum 31 = Includes land areas which are 1% to 29% cultivated at any time during the year, and/or pasture or grazing lands. Minimum block size = 6 km^2. Delineation and shading in *orange*.

Stratum 32 = River beds. These are land areas where the watercourse of a river has changed over time, and which are subject to flooding. Soil patterns are distinct on satellite imagery. Often an embankment is present on both sides of the stratum. Minimum block size = 6 km^2. Delineation is in *orange* and shading in *dark blue*.

Stratum 41 = Land not suitable for cultivation or pasture such as deserts, high mountains, barren escarpments, national parks, glaciers or any other totally barren areas. Minimum block size = 2 km^2. Delineation and shading in *brown*.

Stratum 42 = Land reserved for national security such as military installations, restricted border areas or any other areas where sampling is prohibited. Minimum block size = 2 km^2. Delineation is in *brown* and shading in *bright green.*

Stratum 51 = Urban areas - city centres, commercial centres, office complexes or other areas where no agriculture or livestock is expected. Minimum block size = 0.2 km^2. Delineation and shading in *green.*

Stratum 52 = Villages and other agri-urban areas that may have agriculture or livestock present. This includes the perimeter area of larger cities, villages and concentrations of ten or more residences which may not be associated with a village. Minimum block size = 0.2 km^2. Delineation is in *green* and shading in *orange.*

Stratum 61 = Forests, woods, or any other concentration of trees or timber. Orchards are excluded. Minimum block size = 2 km^2. Delineation is in *green* and shading in *yellow.*

Stratum 62 = Orchards or other concentrations of grown tree crops. Timber plantations are excluded. Minimum block size = 1 km^2. Delineation is in *green* and shading in *bright green.*

Stratum 70 = Water - any body of water completely surrounded by land such as lakes, reservoirs, and large rivers. Small streams are included in the other strata where they flow. Minimum block size = 1 km^2. Delineation and shading is in *blue.*

Minimum block size refers to the smallest portion of a specific stratum that may be delineated. Portions of a stratum smaller than the minimum block size are left within the stratum in which they are embedded.

5. SEGMENT TARGET SIZE BY STRATUM

For each land use stratum, the target segment size was determined taking into account the following points:

- the availability of recognizable physical boundaries;
- a manageable number of holders and fields so that an enumerator could enumerate a segment in one day.

The following segment target size per stratum was adopted:

Stratum	Area (km^2)
10	0.2
20	0.2
31	0.3
32	0.3
41	0.3
51	0.1
52	0.1
61	0.3
62	0.3

6. SAMPLE SIZE AND ALLOCATION

Phase I. For Phase I, sample sizes were determined at the district level taking into consideration the area of the major crops grown in the district and a 5% CV. Optimum sample sizes were estimated based on crop area estimates from the VMS and using estimates of variance for major crops based on Deming's method of sample curves (cf. Deming, 1960).

Phase II. Between-segment variances obtained in the Phase I surveys and district crop statistics from the provincial departments were used to determine the sample size at province level for the Phase II surveys. A non-linear regression model was developed relating to Phase I survey data and proportion of crop area from the VMS. The model used to estimate the between-segment variance at provincial levels was:

$$S_i^2 = a_i \, (p \, (1-p))^b,$$

where, p denotes the proportion of cultivated area planted to a crop in a district, i is the stratum subindex, and a and b are the parameters estimated from the data. Using this model, the estimated sample sizes were optimally allocated to nine provincial-level strata. Then these allocations were proportionally allocated to district-level strata. Finally, these district-level allocations were adjusted so that:

- all districts had at least two segments;
- no expansion factor was increased by more than 10% relative to the corresponding optimal allocation expansion factor; and
- all district strata sample segment allocations were divisible by two so that variance estimates could be calculated in each district stratum.

Non-agricultural stratum allocations were adjusted. The allocation was also adjusted to correspond more closely to the intensity of agriculture in the districts. This finalized the sample allocation at the province/district level.

The distribution of the total sample size (4 278 segments) by province, district and stratum is given in Tables 9.2 to 9.5.

7. CONSTRUCTION OF THE STRATA AND PSUs. SIZE OF PSUs

Since 1989, SPOT satellite images with 20m resolution in scale 1:100 000 and topographic maps in scale 1:50 000 have been used for area frame construction, thus substituting the construction of photomosaics as done in Phase I. District boundaries with recognizable physical marks and PSUs were delineated on the topographic maps.

The SPOT images were printed in the standard false colours. Band 1 is printed in blue, band 2 is printed in green and band 3 in red. Band 2 is best for showing roads and bare soils; it also heightens the contrast between areas with and without vegetation. Band 3 is the near infrared band. It helps evaluating plant matter quantity and separates water bodies from vegetation. Band 3 is not as effective as band 2 for road detection. A manual was developed as an aid in interpreting the colour variations caused by advancing stages of crop development, soil treatment and other land use such as forests, marshes, beaches.

Dividing the land area of a province into sample units (segments) would have been extremely time consuming. To avoid this, each stratum block was divided into PSUs with readily definable boundaries and a target size of ten segments, that could vary between 2 and 15 segments. The PSUs were measured and assigned a number of segments based on the segment size for that particular stratum (cf. Volume I for more detail).

The homogeneous land use areas (strata blocks) were delineated directly on topographic maps with the guide of satellite imagery. Three operational methods were used for working with the satellite images:

1. A direct side-by-side viewing of the satellite imagery and the map.
2. The map was made on a transparency and overlaid on the imagery. Boundaries were drawn on the transparency and later transferred to the basic frame map. This method was used primarily for delineating the strata blocks and some of the PSU boundaries using roads as shown on the maps.
3. The satellite imagery was projected on to the frame map using a zoom transfer scope for the delineation of strata blocks and PSUs. One of the most important advantages of the use of a zoom transfer scope is the ability to record accurately changes directly on to the frame map. Changes in river courses, new canals, new highways, etc. can be placed precisely on the map ensuring accurate measurements of the PSUs.

For the construction of strata and stratum blocks, extensive field verification was undertaken to complement the available cartography (including the satellite imagery).

The measurement of strata, PSUs and other areas was made by using a digitizing table, thus reducing errors and time as compared with Phase I.

8. SAMPLE SELECTION

8.1 Sample Selection for the Entire Country except Special Areas and Terraced and other Problem Areas

As has already been mentioned, this survey area covers more than 90% of the entire national territory.

For the sample selection, the first step was to select a random sample of PSUs for each replicate. These selected PSUs were then divided into segments, using photography, and one segment was chosen at random to be the sample segment. See Volume I for detail on the selection of a replicated sample. PSUs containing more than 15 segments were broken into subdivisions with an equal number of segments. One subdivision was selected at random and was subdivided into segments. Then one segment was selected at random.

8.2 The Point Sampling Method Applied in Special Areas. Construction of Sample Segments

Phase I: In some of the sensitive border areas and in areas where photography was not available, or physical features were rare, it was not possible to subdivide PSUs into segments based on what could be seen on the topographic map, therefore point sampling was used. In these cases, a grid was imposed over the selected PSU and a random point was selected. The selected sample point was established by locating the permanent land marks around it and measuring the distances from identifiable features in the map. To form the segment in the field, the segment centre point was taken to be the nearest field corner from the located random point. In cultivated areas, a circle centred in the selected point, with a radius of 250 m (corresponding to the target segment size), was identified in order to orient the construction of the segment with recognizable physical boundaries. Since the segment boundaries must follow the physical features, the "in-out" rule was adopted for the fields which were intersected by the circle. The segment was defined as the land of all complete fields that had 50% or more inside the circle. A rope with the length of the radius was used to identify the fields to be included in the segment. If more than a half of a cultivated field was included inside the circle, the entire field became part of the segment, otherwise the field was excluded from the segment. If it was not clearly visible, the area was measured. In case of half-half field, the coin was flipped. When large areas of non-agricultural land were partially inside the circle, only the portion inside the circle was included in the segment. A scale drawing of the segment was made as it was being laid out. In intensively cultivated areas, such preparation is very costly and time consuming when compared with the photo enlargements.

Phase II: In Phase II, the selected points marked on the topographic maps were also marked on their duplicates in order to obtain the contact prints. The random points were transferred to the contact prints and a specific area, around these points, was marked and their photo enlargements at 1:5 000 scale were prepared. The sample points were transferred from contact prints to photo enlargements. Circles were drawn around these points on the enlargements. The "in-out" rule was used for boundary fields, as discussed in Phase I. After reviewing and checking, the segment boundaries were marked with red ink and measured.

The construction of sample segments through field operation, i.e. the preparation of overlays improved in Phase II. Two methods were used in locating the selected sample point:

- By identifying reference boundaries (roads, mosques, watercourses, etc.) shown on the topographic maps, as described in Phase I.
- By using a Global Positioning System (GPS) instrument.

Some selected sample points could not be located because of lack of reference features on the topographic maps. A GPS instrument was used for this purpose. The GPS uses the longitude and latitude values with reference to the selected sample point which are available from the topographic maps. The point is identified as closely as the ground features allow. The GPS is turned on and the longitude and latitude values are entered and start moving in the expected direction until the instrument shows "Close" on the screen. Then the operator moves in the direction which reduces the distance to the sample point until the word "Complete" appears on the screen, and stops at this point. This is the required point. The GPS takes its guidance from satellites present in the area at the time and locates the point. Then, to construct the layout, the point is shifted to the nearest corner of the field in which it lies. Further construction was similar to the procedure used in Phase I. Compasses were used in determining the directions while drawing the scaled fields.

This point sampling selection was used in only an estimated 10% or less of the national area.

8.3 Sampling Method Applied in Terraced and other Special Areas. Construction of Sample Segments - Phase III

In one province, the frame was incomplete and the segments were not delineated around the selected sample points. This problem appeared in terraced/mountainous areas.

As already mentioned, the selected segment's boundaries and the boundaries of the fields within these segments were demarcated on the enlarged aerial photography or satellite images, or by manually drawing the area segments on scaled paper. These procedures are more accurate where the areas are on a plane and the fields are of regular shape. In terraced areas, instead, neither photography nor imagery gives the true picture of the physical features and the laying out of the terraced fields cannot be done accurately as the fields are of irregular shape. Aerial photography and/or satellite imagery also gives an underestimation of the area owing to the nature of the slope. The correct measurement of areas according to the scale on the segment photo enlargement was not possible. Preparing the layouts by field operation was even more erroneous. The drawing of terraced walls was another source of error.

A line segment approach was adopted to resolve these problems. The first step was the determination of the optimal length of the line segment, i.e. the length of the line, giving the same variability as obtained through the area segment approach. Data of area segments were synthesized to make a study for the purpose. The line length was determined by using and comparing variance models and analysis of variance techniques. After determining the length for each stratum, the fieldwork started. The sample points were located, with the help of topographic maps/GPS and the lines were drawn North-South, with the help of compasses. Fields crossed by the line were also sketched. The line lengths of the fields were converted into areas, proportionate to the target segment areas. The best possible techniques were adopted to remain in line with the estimation/expansion procedures of area sampling methods. The lengths and procedures will be evaluated at the availability of field data. With the completion of these segments, the survey area frame will almost cover the entire country.

Another problem area is the Thar Desert area, which is not important as far as agriculture is concerned. The major problems are: finding and reaching the segments, providing for travelling arrangements and guidance, and finding the operators. Other problem areas are those in which the flooded rivers cover, every year, the fields within their flood ranges and the segment layouts have to be constructed every year before the survey field data collection.

9. ESTIMATION METHODS

The sample data collected from the segments in each stratum are expanded to the stratum level. The expanded data for each land use stratum can then be summed to arrive at estimates for a province or the entire country. The formulae using replicates for the direct expansion area sample estimators and their variances are presented in Chapter 11.

10. DATA COLLECTION PROCEDURES

There are two major crop seasons in Pakistan, *rabi* and *kharif*, hence two annual survey rounds are conducted, i.e. the Rabi Crops Survey and the Kharif Crops Survey.

During Phase I, questionnaires, survey forms as well as an enumerator manual and other materials for instruction were prepared. In Phase II, the surveys continued in the seven districts together with the construction of a national level sampling frame.

Enumerators from each province are trained before each survey round to collect data from farmers through personal interviews. The enumerators were provided with a photographic enlargement (in scale 1:5 000) or an overlay of an enlargement for the selected segment with its boundaries appropriately marked. A topographic map is also provided to guide the enumerator to the segment in the field. Chapter 8 in Volume I discusses data collection procedures in more detail.

In each province supervisors conduct quality control visits to the same segments visited by the enumerators and make checks to verify whether the farmers were interviewed and the data had been recorded accurately by the enumerators.

Training courses for enumerators are conducted at provincial headquarters shortly before launching each survey round. The same team conducts training at each station. During Phase II, the same procedures continued to be applied. However, in Phase III, the master trainers responsible for conducting provincial training courses had one observer from the central team attending the training course.

11. DATA EDITING, TRANSMISSION AND SUMMARIZATION

Questionnaires were edited/coded by the enumerators and supervisors and sent to the Federal Bureau of Statistics (FBS) headquarters in Islamabad where a computer laboratory was established. Data were key-entered and again checked and cleaned through computer edit checks. A central team was responsible for preparing and disseminating the results. In Phase III, computer laboratories were established at provincial headquarters, where the data were entered and cleaned and sent to the headquarters of the Federal Bureau of Statistics, where information submitted by the provinces was analysed and estimates at the national level were prepared according to a crop calendar schedule.

12. OBJECTIVE CROP YIELD SURVEYS BASED ON SUBSAMPLES OF THE AREA SAMPLE FRAME

The determination of sample sizes was based on five major crops i.e. wheat, cotton, sugar cane, rice and maize. The objective crop yield surveys (OYS) for these five crops were also planned. Studies were undertaken to determine the optimal plot sizes in comparison with the plot sizes already in use by provincial agricultural departments.

In the sample of crop fields selected from the area frame at the district level, plots of particular sizes for each crop were identified. These plots were harvested at their maturity and sent to the headquarters of OYS laboratories for the determination of their outputs. The yields were determined at plot, district, provincial and national levels. In Phase III, the area and objective yield surveys of Phase I were discontinued and substituted by provincial-level area and objective yield survey.

In Phase III, OYS laboratories were established at all provincial headquarters. Only data in diskettes were supplied to headquarters for the preparation of results at the provincial and national levels. At present, the provincial staff are also trained to do these activities at the provincial level.

It was also planned to develop a forecasting model based on the plant characteristics at different stages of maturity. Data were collected during the first two phases and forecasting models were developed for *wheat* and *cotton*. They worked as expected but were supposed to improve with the availability of more data. Unfortunately, at the national level there were some detrimental data gaps for these crop forecasting models.

Table 9.1
Pakistan: wheat area for harvest
Area sampling frame *vis-a-vis* Village master sample
(in 1 000 ha)

		Area sampling frame				Village master sample		
Province	District	89/90	90/91	91/92	92/93	89/90	90/91	91/92
	Faisalabad	266.6	262.4	262.6	259.5	260.0	259.0	260.0
	C.V. (%)	3.6	4.0	3.8	3.8			
	Jhang	344.5	336.7	348.6	363.6	318.0	320.0	322.0
	C.V. (%)	5.7	5.6	5.3	6.0			
	Multan	303.5	339.1	339.0	357.7	311.0	324.0	308.0
	C.V. (%)	6.2	5.7	5.3	5.3			
	Sheikhupura	277.3	266.5	269.0	269.3	256.0	254.0	253.0
	C.V. (%)	4.2	4.4	4.0	4.1			
Punjab		1 191.9	1 204.7	1 219.2	1 250.1	1 145.0	1 157.0	1 143.0
	C.V. (%)	2.6	2.6	2.4	2.6			
	Nawabshah	198.1	193.6	193.6	172.2	199.0	201.0	202.0
	C.V. (%)	5.7	5.8	6.3	7.0			
	Larkana	57.8	46.2	36.3	20.0	55.0	55.0	56.0
	C.V. (%)	14.2	17.1	16.7	19.7			
	Hyderabad	123.8	102.1	99.6	88.0	105.0	105.0	106.0
	C.V. (%)	8.5	8.7	9.6	10.0			
Sindh		379.7	341.9	329.4	280.2	359.0	361.0	364.0
	C.V. (%)	4.6	4.8	5.1	5.5			
Total		**1 571.6**	**1 546.6**	**1 548.6**	**1 530.4**	**1 504.0**	**1 518.0**	**1 507.0**
	C.V. (%)	**2.3**	**2.3**	**2.2**	**2.4**			

Table 9.2

Pakistan
Punjab Province: sample size by district and stratum

District	Strata									
	10	20	31	32	41	51	52	61	62	Total No. of segments
Attock	28	12	14	2	4	2	2	6	0	70
Bahawalnagar	60	12	6	2	4	2	2	2	0	90
Bahawalpur	48	8	6	2	10	2	2	2	0	80
Bhakkar	42	4	8	0	2	2	2	2	0	62
Charkal	34	10	6	0	4	2	2	2	0	60
D.G. Khan	30	6	6	4	10	2	2	2	0	62
Faisalabad	70	4	2	2	0	2	2	2	0	84
Gujranwala	64	4	2	2	2	2	2	0	0	78
Gujrat	52	6	4	2	2	2	2	2	2	74
Islamabad	2	6	2	0	2	2	2	2	0	18
Jhang	74	12	4	4	2	2	2	2	0	102
Jehlum	14	2	12	0	2	2	2	4	0	38
Kasur	40	2	4	2	0	2	2	2	0	54
Khanewal	32	2	2	2	2	2	2	2	2	48
Khushab	54	4	6	0	4	2	2	2	0	74
Lahore	16	2	0	2	0	2	2	2	0	26
Leiah	38	12	4	4	4	2	2	2	0	68
Mianwali	48	6	10	4	4	2	2	2	2	80
Multan	86	6	2	2	0	2	2	2	0	102
Muzaffargarh	50	10	6	6	4	2	2	4	2	86
Okara	46	2	2	2	0	2	2	2	2	60
Rahimyarkhan	76	4	2	2	8	2	2	2	2	100
Rajanpur	34	2	2	4	6	2	2	2	0	54
Rawalpindi	12	10	18	0	2	2	2	4	0	50
Sahiwal	68	2	2	0	2	2	2	2	0	80
Sargodha	62	2	2	2	2	2	2	2	2	78
Sheikhpura	64	4	2	2	2	2	2	2	0	80
Sialkot	58	2	2	2	2	2	2	2	0	72
Toba Tek Sing	28	4	2	0	0	2	2	2	0	40
Vehari	48	0	2	2	2	2	2	0	0	58
Total	**1 378**	**162**	**142**	**58**	**88**	**60**	**60**	**66**	**14**	**2 028**

Table 9.3
Pakistan
Sindh Province: sample size by district and stratum

District	Strata									
	10	20	31	32	41	51	52	61	62	Total No. of segments
Badin	80	8	2	0	2	2	2	2	0	98
Dadu	38	8	8	2	14	2	2	4	2	80
Hyderabad	68	2	2	2	2	2	2	4	0	84
Jacobabad	48	12	6	2	2	2	2	2	0	76
Karachi	2	4	2	0	4	2	2	0	0	16
Khairpur	38	10	4	4	12	2	2	2	0	74
Larkana	52	10	4	2	2	2	2	2	0	76
Nawabshah	66	12	2	2	2	2	2	8	0	96
Sanghar	66	6	4	0	6	2	2	2	0	88
Shakarpur	28	4	4	4	0	2	2	4	0	48
Sukkur	54	6	4	2	6	2	2	6	2	84
Tharparkar	86	10	68	0	10	2	2	2	0	180
Thatta	36	24	16	2	8	2	2	6	0	96
Total	**662**	**116**	**126**	**22**	**70**	**26**	**26**	**44**	**4**	**1 096**

Table 9.4
Pakistan
North-West Frontier Province: sample size by district and stratum

District	Strata									
	10	20	31	32	41	51	52	61	62	Total No. of segments
Abbotabad	8	2	26	2	2	2	2	6	0	50
Bannu	44	12	8	2	2	2	2	0	0	72
Charsada	16	2	2	0	2	2	2	0	0	26
Chitral	4	2	2	0	18	2	2	6	0	36
Dir	10	4	6	0	4	2	2	18	0	46
D.I. Khan	58	32	16	2	8	2	2	2	0	122
Karak	10	4	6	0	2	0	2	0	0	24
Kohat	10	6	8	0	4	2	2	2	0	34
Kohistan	2	2	4	0	8	0	0	10	0	26
Malakand	6	2	0	0	2	2	2	0	0	14
Mansehra	12	10	16	0	4	2	2	16	0	62
Mardan	30	2	4	2	2	2	2	2	0	46
Peshawar	28	2	4	2	2	2	2	2	0	44
Sawabi	6	0	2	2	2	2	2	0	0	16
Swat	28	8	8	0	8	2	2	16	2	74
Total	**272**	**90**	**112**	**12**	**70**	**26**	**28**	**80**	**2**	**692**

Table 9.5

Pakistan
Balochistan Province: sample size by district and stratum

District	Strata									
	10	20	31	32	41	51	52	61	62	Total No. of segments
Chaggi	2	4	3	0	0	1	0	0	0	10
Dera Bughtti	4	5	8	0	0	0	0	0	4	21
Gwadar	3	2	2	0	0	2	0	0	0	9
Cachhi	12	7	5	0	0	1	0	0	0	25
Kalat	13	18	5	0	0	0	0	0	16	52
Kharan	3	6	4	0	0	0	0	0	0	13
Khuzdar	3	4	14	0	0	2	0	0	0	23
Lasbela	14	10	5	0	0	0	0	0	0	29
Loralai	7	12	16	0	0	2	2	0	11	50
Nasirabad	76	22	6	0	0	0	0	0	0	104
Panjgur	0	0	0	0	0	0	1	0	2	3
Pashin	5	12	12	0	0	1	4	0	20	54
Quetta	0	4	6	0	0	6	1	0	4	21
Sibbi	2	3	2	0	0	1	0	0	2	10
Turbat	0	0	2	0	0	0	2	0	6	10
Zhob	3	6	10	0	0	0	2	0	7	28
Total	**147**	**115**	**100**	**0**	**0**	**16**	**12**	**0**	**72**	**462**

CHAPTER 10

COLOMBIA, COSTA RICA, THE DOMINICAN REPUBLIC, EL SALVADOR, GUATEMALA, HONDURAS AND NICARAGUA

GENERAL PURPOSE AGRICULTURAL SURVEY PROGRAMMES BASED ON AREA FRAME SAMPLING METHODS

This chapter describes some characteristics of the area sample survey programmes, and their area frames, which were carried out in the Central American countries, in Colombia and the Dominican Republic, and which, in a few instances, are still used. The current Multiple Frame Agricultural Survey Programme in Honduras and the current Area Frame Agricultural Survey Programme in Nicaragua have been described in separate chapters.

The purpose of this summary is to illustrate further the characteristics of the area survey designs in these countries as possible stratum definitions, size and tolerance limits for counting units and segments and to show sample sizes with replicates.

The National Agricultural Survey Programmes described for the five Central American countries (Costa Rica, El Salvador, Guatemala, Honduras and Nicaragua) and the Dominican Republic have been operational until several countries of the region were hampered by political instability. The sample designs are considered, even today, to be very useful for the establishment of new Multiple Frame Agricultural Survey Programmes in those countries.

The Survey Programmes for the Central American countries are covered in the first section. The National Agricultural Survey Programme of Colombia, currently being redesigned, and that of the Dominican Republic are covered in sections 2 and 3.

1. AREA FRAME AGRICULTURAL SURVEY PROGRAMMES IN CENTRAL AMERICAN COUNTRIES (COSTA RICA, EL SALVADOR, GUATEMALA, HONDURAS AND NICARAGUA)

The area frame for the Dominican Republic was constructed in Washington D.C., at NASS (then the Statistical Reporting Service) in 1970. The maps and aerial photography were sent to the Secretary of Agriculture in Santo Domingo, and the first survey round, which was primarily to estimate coffee production, was completed in 1971.

The area frames in Costa Rica, El Salvador, Guatemala, Honduras and Nicaragua were initiated and constructed during the period 1974-1982.

Political instability, lack of technical assistance and budget have kept the frames from being maintained and updated. The frames in Nicaragua and Guatemala appear to be lost. Costa Rica has made intermittent use of its area sample, conducting national-level surveys in 1982, 1986, 1987, 1988 and 1989. In El Salvador, the condition of the frame following the long period of political instability is such that it cannot be used effectively. In Honduras, the Agricultural Survey Programme has been conducted regularily twice a year for basic grain surveys, but the area frame has not been updated.

1.1 Stratum Definitions, Counting Units and Segments

Table 10.1 shows the stratum descriptions, counting unit sizes and segment sizes for the Central American countries. In general, the strata numbering proceeds from the most intensive agricultural use, to extensive agricultural use, to non-agricultural use, except in the case of El Salvador where Stratum 8 with some cultivation and pasture appears after Strata 6 and 7 where no agriculture is expected.

Where it seemed desirable to form substrata, the numbering sequence followed was 10, 20, 30, ..., 80: from most intensive, to extensive, to no agriculture. Substrata in the most intensive stratum were numbered 11, 12, 13; substrata in the next less intensive stratum were numbered 21, 22, 23, etc.

El Salvador. Originally there were 11 strata. Stratum 11 was projected water where land disappeared when inundated by a hydroelectric project.

Guatemala. In practice, Stratum 6 was divided into three substrata: 61 = the capital city, 62 = other cities, and 63 = agri-urban. The CU sizes for Stratum 6 are shown below in city blocks.

Substrata	*Minimum*	*Optimum*	*Maximum*
61	1	20	40
62	1	10	20
63	2	6	14

Honduras. Stratum 1 is divided into three substrata: 11 = areas without coffee or irrigation; 12 = areas with more than 30% of the cultivated area in coffee; and 13 = land with irrigation. Stratum 6 is divided into two substrata: 61 = urban; and 62 = agri-urban. In Stratum 6, the minimum CU size is six blocks or 50 houses or more in 0.5 km^2; the optimum size is six to ten city blocks.

Nicaragua. The city stratum (6) is divided into urban and agri-urban. Segments were formed within the minimum and maximum size range (cf. Table 10.1). Each segment had its own individual expansion factor.

1.2 The Area Samples

Costa Rica. Table 10.3 shows the final sample proposed for the national survey. It was selected by stratum within regions.

El Salvador. The sample whose parameters are shown in Table 10.4 was designed to provide regional estimates and had been performing well until the country was overtaken by political instability.

Guatemala. No information is available concerning the national sample design of Guatemala.

Honduras. Chapter 3 in this volume refers to the national sample used.

Nicaragua. Table 10.2 shows the original sample prepared for Nicaragua circa 1975. The sample was selected by stratum within regions. Where the population and sample were small, as in Strata 4 and 5, two replicates were selected. Four replicates were selected in Strata 1, 2 and 3. The new frame is presented in Chapter 8.

Table 10.1

Area frame agricultural survey programmes in Costa Rica, El Salvador, Guatemala, Honduras and Nicaragua.
Stratum descriptions, counting unit and segment sizes by country

Country	Stratum Description	Counting Unit Size (Km^2)			Segment Size (Km^2)		
	STRATUM 1	Min.	Opt.	Max.	Min.	Opt.	Max.
Costa Rica	60-100% cultivated	1	4-6	9	0.25	0.50	0.75
El Salvador	75-100% cultivated	2	10	18	0.25	0.50	0.75
Guatemala	60-100% cultivated	2	10	20	0.75	1.00	1.25
Honduras	60-100% cultivated	1	4-6	8	0.25	0.50	0.75
Nicaragua	60-100% cultivated	2	8-12	18	0.50	---	1.00
	STRATUM 2						
Costa Rica	30-59% cultivated	2	7-11	16	0.75	1.00	1.25
El Salvador	50-74% cultivated	2	10	18	0.25	0.50	0.75
Guatemala	30-59% cultivated	4	16	32	1.00	2.00	3.00
Honduras	30-59% cultivated	2	8-12	16	0.75	1.00	1.25
Nicaragua	25-59% cultivated	2	8-12	18	0.50	---	1.00
	STRATUM 3						
Costa Rica	0-29% cultivated	6	22-32	48	0.50	1.50	2.50
El Salvador	25-74% cultivated with at least 50% of cultivated land in coffee	2	10	18	0.25	0.50	0.75
Guatemala	0-29% cultivated	8	24	40	2.00	4.00	6.00
Honduras	0-29% cultivated	4	20-30	40	1.50	2.00	2.50
Nicaragua	0-24% cultivated	4	12-20	28	1.00	---	2.00
	STRATUM 4						
Costa Rica	Not suitable for crops or livestock	8	30-42	64	Variable		
El Salvador	15 - 49% cultivated	4	16	28	0.75	1.00	1.25
Guatemala	Suitable for cultivation but not being used	8	28	48	2.00	4.00	6.00
Honduras	Suitable for cultivation but not being used	8	32-48	60	3.00	4.00	5.00
Nicaragua	Not suitable for crops or livestock	1	Open		Variable		

Table 10.1 (Continued)

Area frame agricultural survey programmes in Costa Rica, El Salvador, Guatemala, Honduras and Nicaragua.
Stratum descriptions, counting unit and segment sizes by country

Country	Stratum Description	Counting Unit Size (Km^2)			Segment Size (Km^2)		
	STRATUM 5	Min.	Opt.	Max.	Min.	Opt.	Max.
Costa Rica	Cities and towns	0.5	2-6	8	---	0.01	---
El Salvador	Ag-urban (some agriculture present)	0.5	4	8	---	0.25	---
Guatemala	Not suitable for cultivation	2	10	30	1.00	2.00	3.00
Honduras	Not usable for crops or livestock	4	20-30	40	1.50	2.00	2.50
Nicaragua	Cities and towns	10 city blocks			one city block		
	STRATUM 6						
Costa Rica	Ag-urban (some agriculture present)	0.5	2-6	8	---	0.02	---
El Salvador	Urban areas (no agriculture)	0.5	4	8	---	0.10	---
Guatemala	Populated areas	see 1.1			one city block		
Honduras	Cities and towns	see 1.1			one city block		
Nicaragua	Large bodies of water	1	open		not sampled		
	STRATUM 7						
Costa Rica	Large bodies of water	0.5	any size		not sampled		
El Salvador	Tourist and recreation areas	0.5	4	8	0.25	0.50	0.75
Guatemala	Large bodies of water	2	open		not sampled		
Honduras	Large bodies of water	1	open		not sampled		
	STRATUM 8						
El Salvador	0 - 14% cultivated	4	30	56	1.00	2.00	3.00
Honduras	Projected water	sizes for strata 1, 2 and 3 as appropriate			optimum range = 0.50 to 2.00		
	STRATUM 9						
El Salvador	Not usable for crops or livestock	2	16	30	1.00	2.00	3.00
	STRATUM 10						
El Salvador	Large bodies of water	1	open		not sampled		

Table 10.2

NICARAGUA: national area sample

Stratum	Pacific Region	Central Region	Total
1 or 10	102	81	183
2 or 20	34	61	95
3 or 30	17	41	58
4 and 5 or 40 and 50	17	20	37
Totals	**170**	**203**	**373**

Table 10.3

COSTA RICA: national area sampling frame and area sample

Region	Stratum	Segment Population (N)	Replicates (R)	Segment per Replicate	Total Sample (n_h)
Chorotega (# 1)	I	2 316	8	6	48
	II	320	6	4	24
	III	11 440	6	8	48
	IV	647	1*	2	2
	VI	1 159	1*	4	4
					n_1 = 126
Central (# 2)	I	3 276	5	14	70
	II	58	1*	4	4
	III	9 324	6	6	36
	IV	103	1*	2	2
	V	1 463	1*	4	4
	VI	8 395	6	5	30
					n_2 = 146
Brunca (# 3)	I	1 224	5	8	40
	II	116	1*	8	8
	III	6 366	4	6	24
	IV	156	1*	2	2
	VI	417	1*	4	4
					n_3 = 78
Huetar - Atlantica (# 4)	I	1 015	5	7	35
	II	240	4	4	16
	III	11 830	7	7	49
	IV	809	1*	4	4
	VI	343	1*	4	4
					n_4 = 108
Huetar - Norte (# 5)	II	40	1*	3	3
	III	5 784	4	6	24
	VI	156	1*	2	2
					n_5 = 29
Costa Rica					n = 487

Note: 1* indicates a simple random sample, with no replicates.

Table 10.4
EL SALVADOR: national area sampling frame and area sample

Region	Stratum	Segment Population (N_h)	Replicates (R)	Segments per Replicate	Total Sample (n_h)
Region I - West	1	1 561	4	7	28
	2	910	3	5	15
	3	2 556	4	9	36
	4	834	2	6	12
	5	115	1	5	5
	6	100	1	4	4
	7	14	1	2	2
	8	770	2	5	10
	9	14	2	2	4
		N_I = 6 874			n_I = 116
Region II - Central	1	1 608	5	6	30
	2	930	4	5	20
	3	1 935	4	5	20
	4	1 500	4	6	24
	5	318	2	6	12
	6	258	2	6	12
	7	42	2	3	6
	8	1 440	4	5	20
		N_{II} = 8 031			n_{II} = 144
Region III - Central	1	2 514	5	6	30
	2	1 710	4	5	20
	3	408	2	6	12
	4	546	4	6	24
	5	80	2	2	4
	6	20	2	2	4
	7	36	2	3	6
	8	540	4	5	20
	9	36	1	2	2
		N_{III} = 5 890			n_{III} = 122
Region IV - East	1	2 436	5	6	30
	2	3 420	5	10	50
	3	1 310	2	10	20
	4	2 124	5	6	30
	5	240	2	5	10
	6	60	1	3	3
	7	16	1	2	2
	8	1 495	4	5	20
	9	177	2	3	6
		N_{IV} = 11 278			n_{IV} = 171
El Salvador		N = 32 073			n = 553

2. COLOMBIA: THE AREA FRAME SURVEY PROGRAMME

Five pilot surveys testing area frame methods were carried out in Colombia from 1983 through 1986.

An area frame was constructed for the 23 departments considered to be the most important agriculturally and the first survey of selected departments was conducted in 1988. Survey results were published in August 1990. The area frame was constructed using land use maps, aerial photography and topographic maps. Delimitation of counting units and stratification were done simultaneously, i.e. a counting unit was formed and then classified, numbered, etc. according to the stratum definitions (cf. Table 10.5).

Unfortunately, carrying out this survey did not result in the establishment of a permanent agricultural information system. However, in 1991, the Secretary of Agriculture of the department of Antioquia used the area frame to establish a system of Municipal Agricultural Statistics. At the end of 1992, seven departments in western Colombia joined with Antioquia in implementing a system that provides data at the municipal level.

In 1993, the National Department of Planning, the Ministry of Agriculture and the Administrative Department of National Statistics (DANE) agreed that DANE should be responsible for carrying out a periodic Survey Programme to estimate the important agricultural variables. It was decided that the area frame constructed for the 1988 survey would form the base for the new frame and sample. This original frame was enlarged to include the entire country and some changes were made in the stratification from information gathered during the 1988 survey.

The first survey round, using a national sample from the new frame, was conducted during the period April-June of 1995. A second survey round took place in October-December of the same year. The traditionally non-agricultural departments of Amazonas, Vaupes, Guainia and Vichada were not included in the second round of the survey. The 1995 surveys were general purpose, gathering data on total land in farm, land use, area and production of principal annual and permanent crops, livestock inventories, milk production and poultry.

The area sample was selected systematically from the listing of counting units, without replicates (cf. Table 10.6). Figures 10.1 to 10.4 show the area frame construction and sample selection.

DANE is continuing with the system of semi-annual surveys started in 1995, and is at present undertaking studies to redesign the Agricultural Survey Programme.

3. DOMINICAN REPUBLIC: THE AREA FRAME AGRICULTURAL SURVEY

The stratum descriptions and Table 10.7 show the data for the original sample in the Dominican Republic. Segment sizes varied from 0.5 km^2 to 4 km^2.

Stratum I	Area where most of the land is cultivated.
Stratum II	Land sown to crops on the mountainsides at 100-m. elevation or higher - mainly coffee and cacao.
Stratum III	Areas around the cultivated zones that could be used for other crops and/or livestock.
Stratum IV	Marginal land.
Stratum V	Cities and densely populated areas.

Following the first round of the survey, the land in Stratum III was reviewed and substrata were formed according to their similarity to the other strata i.e. Substratum III-1 was most similar to Stratum I; Substratum III-2 had characteristics similar to Stratum II, and so on. The sample was modified as shown in Table 10.8.

After further experience, the technicians in the Dominican Republic modified the original stratification in two provinces as indicated in Tables 10.9 and 10.10.

Table 10.5

COLOMBIA: stratum descriptions with counting unit and segment sizes

Stratum	Sub-stratum	Description and/or definition	Counting unit size (km^2)	Segment size (km^2)
10		Land 70-100% cultivated	See footnote	
	11	Transitory crops in medium and large fields	r = 4-6 opt = 5	r = 0.9-1.1 opt = 1
	12	Permanent crops in medium and large fields	r = 8-12 opt = 10	r = 1.9-2.1 opt = 2
	13	Areas where the primary crop is coffee	r = 9-15 opt = 12	r = 2.9-3.1 opt = 3
	14	Areas where small fields are dominant		
	15	Others		
20		Land 30-69% cultivated	Same as Stratum 10	
	21-25	Same subcategories as Stratum 10		
30		Land 10-29% cultivated		
	31-35	Same subcategories as Stratum 10		
40		Primarily pasture or rangeland: may have up to 9% cultivation	r = 9-15 opt = 12	r = 2.9-3.1 opt = 3
	41	Areas of improved pasture and rangeland		
	42	Brush land, abandoned cropland, etc.		
	43	Rough rangeland, high altitude vegetation		
50		Land covered 90-100% by natural forests or planted woodlands.		
	51	100% forested or wooded land	Not sampled	
	52	90-99% forested or wooded land	r = 8-16 opt = 12	r = 3.9-4.1 opt = 4
60		Areas 90-100% useless for crops or livestock		
	61	Areas 100% useless for crops or livestock	Not sampled	
	62	Areas with up to 9% cropland or pasture	r = 8-16 opt = 12	r = 3.9-4.1 opt = 4
	63	Bodies of water, swamps, large rivers, etc.	Not sampled	
	64	Other	??	??
70		Areas with high population density		
	71	Urban - departmental capitals, city centres, etc.	Not sampled	
	72	Agri-urban - with small amounts of agriculture	0.1	0.01
80		Special Areas - 90-100% cultivated where particular crops are concentrated that are not grown in other parts of the department	Determined according to characteristics of each situation	
90		List frame - very large operators whose land is marked on the area frame materials but not placed in any of the sampling strata. List is basis for data collection	Sampled or enumerated from list	

Note: The ranges (r) and optimum sizes (opt) for CUs and segments in Stratum 10 were applied depending on the number of interviews expected, i.e. larger sizes used where fields were large, smaller sizes where fields were small but only one set of sizes used within a substratum.

Table 10.6

COLOMBIA: national area sample by departments

Departments	Sample Segments (n)
1. Antioquia	500
2. Atlántico	80
3. Bolívar	250
4. Boyacá	300
5. Caldas	150
6. Caquetá	100
7. Cauca	200
8. Cesar	250
9. Córdoba	250
10. Cundinamarca	300
11. Chocó	80
12. Huila	250
13. La Guajira	150
14. Magdalena	250
15. Meta	200
16. Nariño	200
17. Norte de Santander	250
18. Quindío	80
19. Risaralda	100
20. Santander	300
21. Sucre	150
22. Tolima	250
23. Valle del Cauca	380
TOTAL	**5 020**

Table 10.7

THE DOMINICAN REPUBLIC: original national area sample

Stratum	Area (Km^2)	Segments assigned	Replicates selected	Segments per replicate	Total sample
I	7 174	3 191	9	20	180
II	1 808	978	3	20	60
III	29 266	7 873	12	10	120
IV	10 680	2 784	2	10	20
V	109	258	2	10	20
Totals	**49 037**	**15 084**			**400**

Table 10.8

THE DOMINICAN REPUBLIC: revised national area frame and area sample

Stratum	Replicates selected	Segments per replicate	Total sample	Expansion factor
I	9	20	180	17.73
II	3	20	60	16.30
III-1 and 2	4	20	80	21.25
III-3	3	16	48	72.00
III-4	3	10	30	73.33
III-5	2	5	10	14.50
IV	2	10	20	139.20
V	2	10	20	13.05
			n = 448	

Table 10.9

THE DOMINICAN REPUBLIC: revised stratification for two provinces

New strata	Land use and/or crop grown	Percent-age cult-ivated	Definition	Original strata
11	Rice	50-100%	50%+ in rice	I and III-1
12	Coffee	50-100%	50%+ in coffee	II
13	Cacao	50-100%	50%+ in cacao	I, II, III-1
14	Intense cultivation	50-100%	No individual crop with 50% of area	I and/or III-1
15	Sugar cane	50-100%	50%+ in sugar cane	I and/or III-1
30	Extensive cultivation and pasture	15-50%	Mixed agriculture and pasture	III-3
40	Non-agric. land	0-15%	Non-agric. without forest	IV and/or III-4
43	Non-agric. land	0-15%	Non-agric. with forest	IV and/or III-4
50	Urban areas	0-15%	Densely populated	V, III-5

Table 10.10

THE DOMINICAN REPUBLIC: revised frame and sample for two provinces

Strata	N_h	Replicates	Segments per replicate	n_h
11	135	3	3	9
12, 13	198	3	5	15
14, 15	364	4	6	24
30	198	3	6	18
40, 43	400	2	4	8
50	15	1	2	2
TOTALS	**1,310**			**76**

DEPARTAMENTO ADMINISTRATIVO NACIONAL DE ESTADISTICA
DIRECCION SINT - PROYECTO SISAC
COMPONENTE GEOESTADISTICO

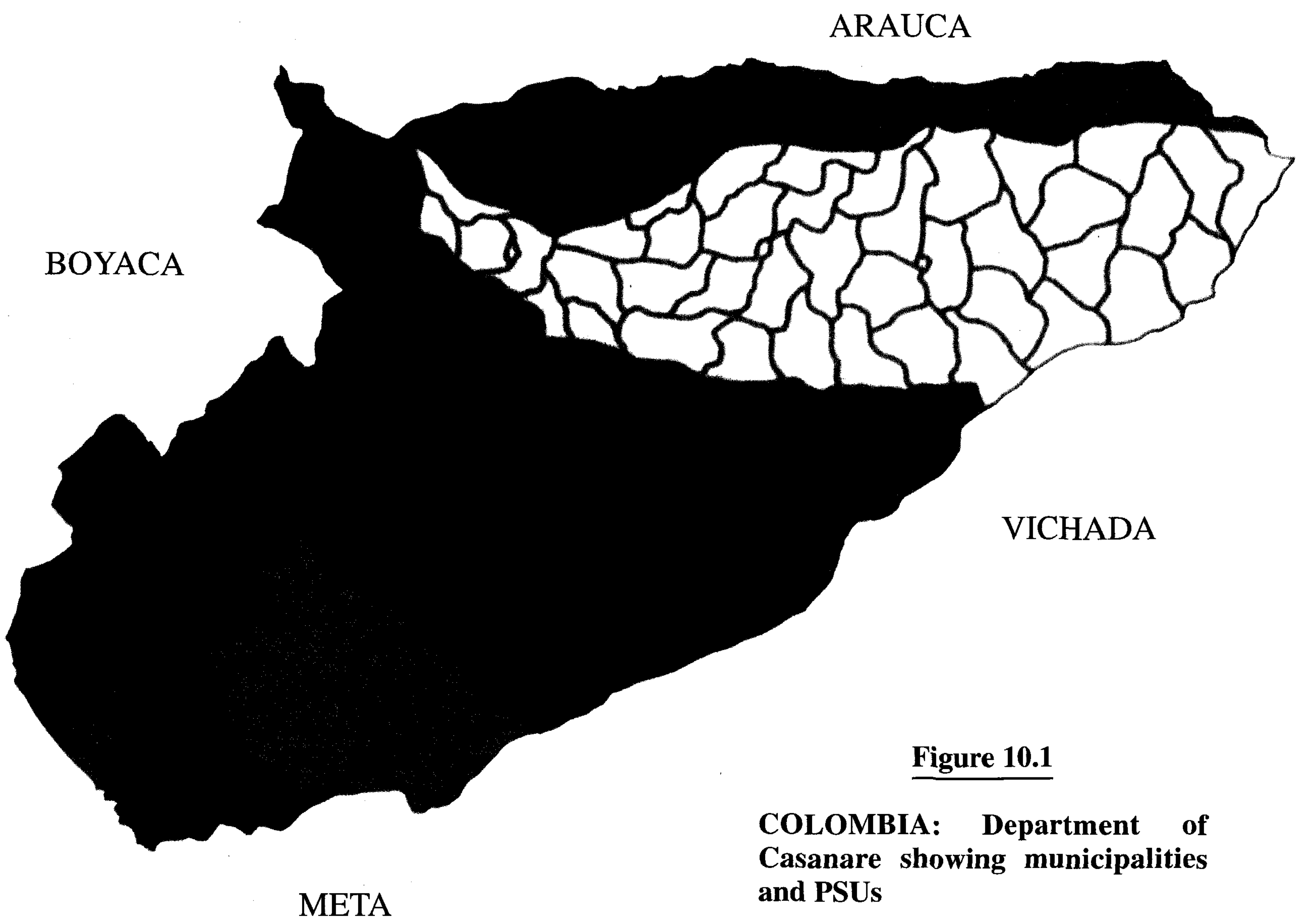


Figure 10.1

COLOMBIA: Department of Casanare showing municipalities and PSUs

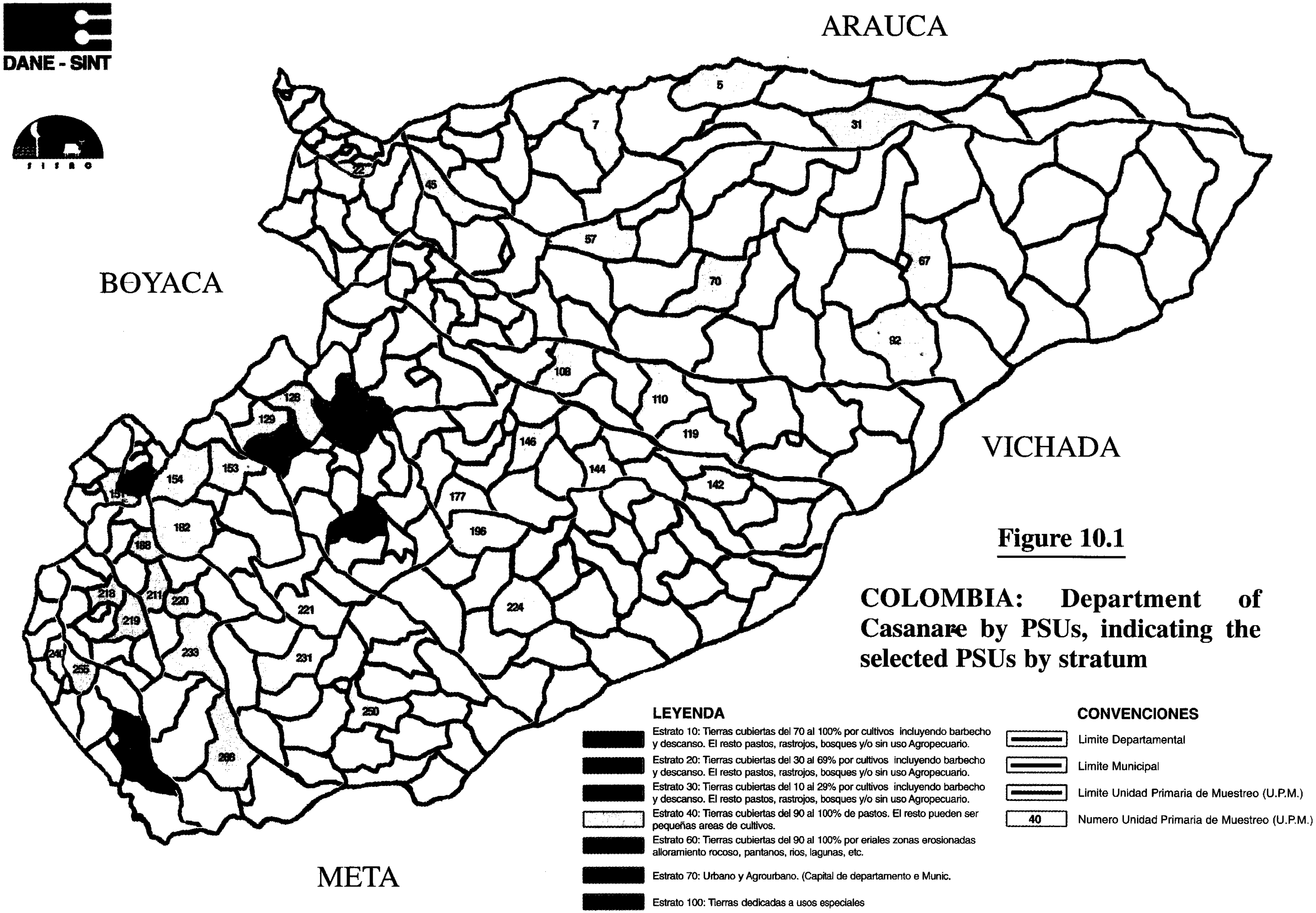
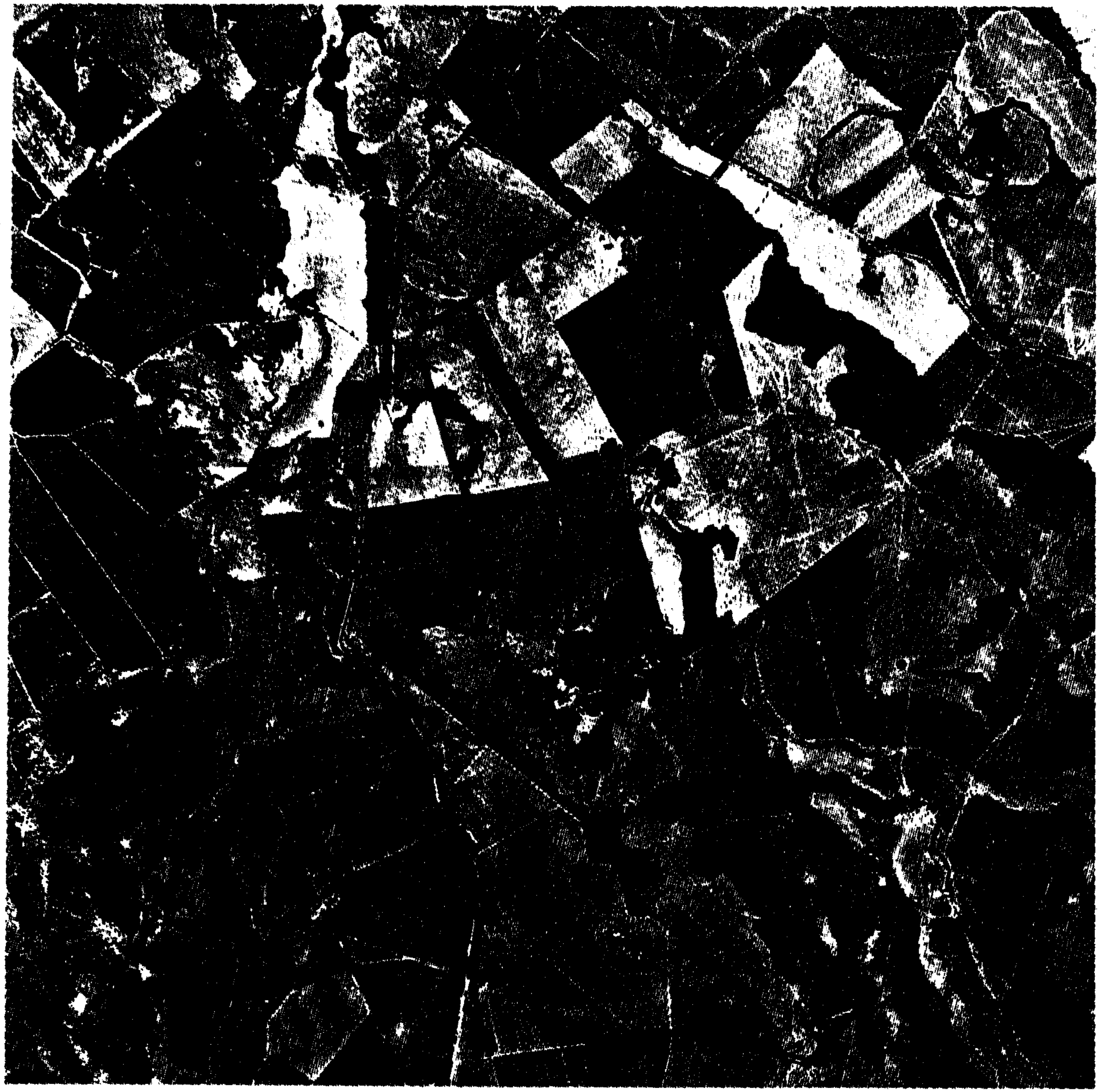


Figure 10.1

COLOMBIA: Department of Casanare by PSUs, indicating the selected PSUs by stratum

DEPARTAMENTO ADMINISTRATIVO NACIONAL DE ESTADISTICA
DIRECCION SINT - PROYECTO SISAC
COMPONENTE GEOESTADISTICO



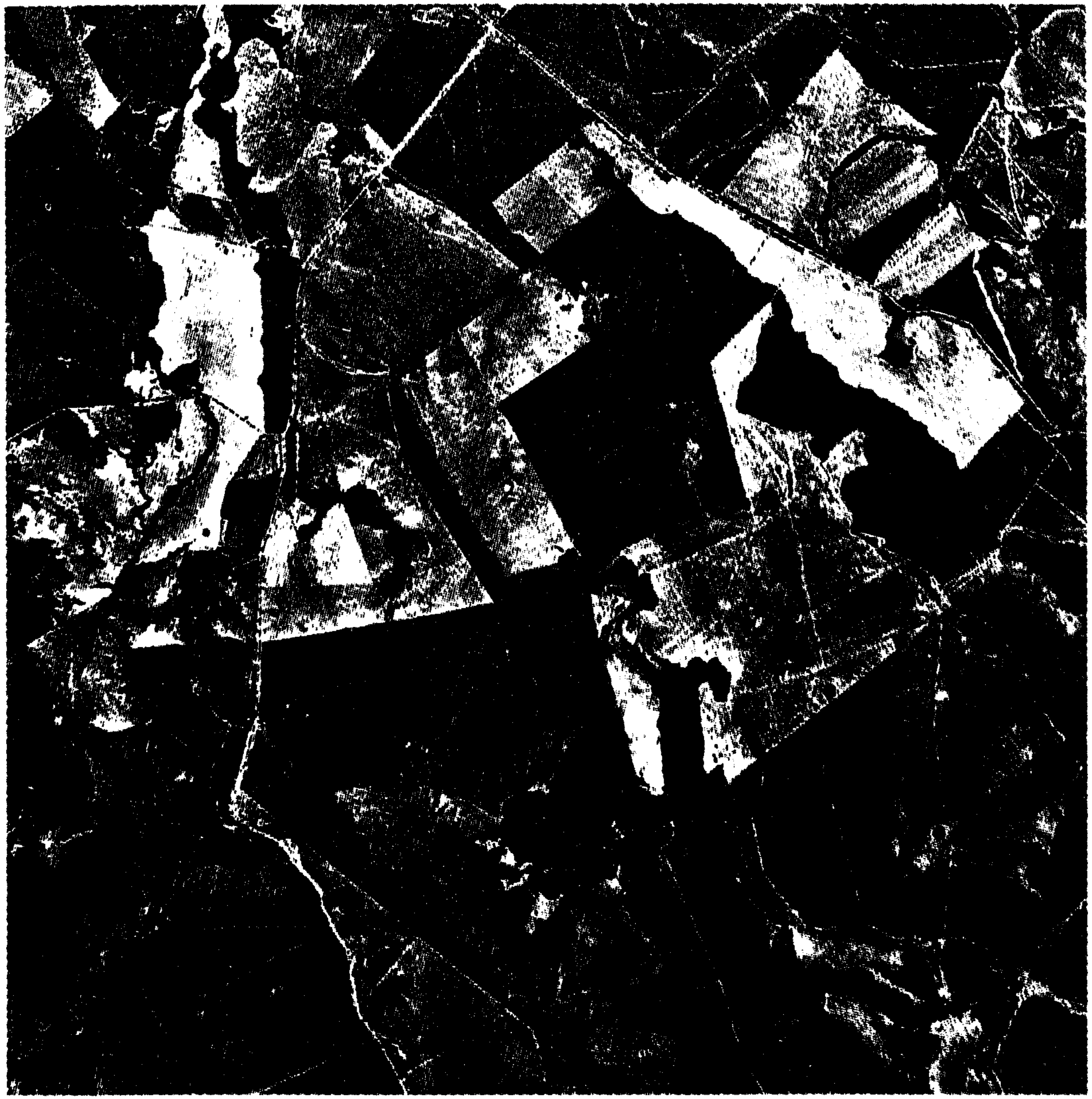
DPTO : CASANARE
MPIO : NUNCHIA
REG : UNICA
E : 10
UPM : 1154
SMNo.1 : 297.17 has
SMNo.2 : 295.15 has
SMNo.3 : 298.76 has
SMNo.4 : 291.33 has
SMNo.5 : 299.04 has
ESCALA : 50.000
FOTO No. : C2563-098
FECHA : Febrero 1995
PLANCHA : 194 III A

ESCALA 1: 50.000
1 0 1 Kms

Figure 10.3

COLOMBIA: Selected PSUs divided into segments

DEPARTAMENTO ADMINISTRATIVO NACIONAL DE ESTADISTICA
DIRECCION SINT - PROYECTO SISAC
COMPONENTE GEOESTADISTICO



DPTO : CASANARE
MPIO : NUNCHIA
REG : UNICA
E : 10
UPM : 1154
SM : 2
ESCALA : 10.000
FOTO No. : C2563-098
FECHA : Febrero 1995
PLANCHA : 194 III A

OBSERVACIONES

AREA DEL SEGMENTO = 295.15 HAS

Figure 10.4

COLOMBIA: Selected segment within a selected PSU

PART II

DIRECT EXPANSION AREA SAMPLE ESTIMATORS AND THEIR VARIANCES, MOST COMMONLY APPLIED IN GENERAL PURPOSE AGRICULTURAL SURVEY DESIGNS

Part II provides a detailed presentation of the direct expansion area sample estimators and their variances most commonly applied in general purpose agricultural survey designs and described in Part I. It also includes the interpretation of the different area sample models in which the formulae are used, as well as dual frame estimators for which a complete enumeration list frame complements the area sample. It has additional considerations and formulae to those given in Chapter 4 of Volume I. Other area sample estimators are included in the appropriate chapters.

CHAPTER 11

DIRECT EXPANSION AREA SAMPLE FRAME ESTIMATORS AND THEIR VARIANCES. INTERPRETATION OF THE AREA SAMPLING DESIGNS AND DUAL FRAME ESTIMATORS

1. INTRODUCTION

This chapter presents the estimation methods for the area sample designs described in Part I, in which the sampling units (segments) have recognizable physical boundaries. These are the most commonly used area sample methods currently applied for general purpose Agricultural Survey Programmes.

The chapter also includes the following:

a) The formulae for the direct expansion and variance calculations for the closed, weighted and open segment estimators, based on substrata or replicates, with the interpretation of the different area sample models in which the formulae are used.

b) Ratio estimators and their variance formulae, both for within-survey and between-survey rounds, for estimators based on substrata or replicates.

c) The multiple frame estimation formulae and their variance for the dual frame designs described in Part I, in which a complete enumeration list frame complements the area sample.

2. THE AREA SAMPLE DESIGNS CONSIDERED

The area sample designs considered consist of a one- or two-stage stratified probability sample of *segments*, with a replicated or non-replicated random or systematic selection procedure.

The area sampling frame considers the survey area divided into a number of land use strata defined by proportion of cultivated land, predominance of certain crops or other land use characteristics. The strata, PSUs and sample segments have *identifiable physical boundaries* (roads, paths, rivers, etc.) that can be located both in the field and on the cartographic materials used for their identification (satellite images, mosaics of aerial photos and maps).

For sample selection, each land use stratum is completely subdivided into a non-overlapping set of PSUs, with a geographic order. Then each land use stratum is completely subdivided, by similarity of agricultural characteristics or following a geographic distribution criterion, into areas with an equal number of segments called *substrata* or *zones*, which provide a further level of stratification.Within each land use stratum, the determination of zones, is done by following the established order of the PSUs. For that purpose, in each land use stratum a list is prepared with the accumulated size measures of the ordered PSUs.

For replicated designs, in each stratum the area sample consists of a number of independent replicates and each replicate is formed by one segment in each substratum, with equal probability of selection within the stratum.

The area sample designs can also be considered as a stratified, cluster sample of *tracts*, a tract consisting of the part of a holding (or non-agricultural areas) included in the segment.

Interpretation of the Sample Designs - Justification for Applying the Estimation Formulae

For the different area sample designs considered (Albania, Argentina, Brazil, Colombia, Costa Rica, the Dominican Republic, El Salvador, Guatemala, Honduras, Morocco, Pakistan and the United States), in order to justify the application of the estimation formulae presented in the following sections properly, it is convenient to consider the following features of the sampling and the estimation methods.

One-Stage vis-à-vis *Two-Stage sampling designs*

In Brazil, a systematic one-stage sample procedure was applied. In fact, to each CU a size measure was assigned proportional to its area (surface) measurement. The CUs were then ordered and a systematic selection procedure was used to identify the CUs corresponding to the selected sample segments. Moreover, a rule was establish to *order* all segments within each CU. As a consequence, once a selected CU was identified, the sample segment contained in the CU was automatically selected. In other words, since the CUs were ordered and a rule was established to order the segments within each CU, the area sampling frame consisted of an *ordered list of segments*. A systematic selection by substrata of this ordered list provided the sample of segments. The estimation formulae applied correspond to a one-stage design and consider a successive difference variance estimation method.

In all other countries, the sampling designs involve a two-stage selection procedure. In fact, the PSUs were also ordered and selected with probability proportional to their size measures, and a segment within each PSU was selected with equal probability (EPSEM). In these cases, each selected PSU had to be subdivided into the assigned number of segments prior to the selection of one segment of the selected PSU. The estimation formulae applied correspond, however, to one-stage sample designs for simplification purposes. It should be noted that the total number of segments is exactly known for each stratum and PSU, and the probability of selection of each segment (constant within a stratum) is equal to its conditional probability in two-stage sampling.

Replicated vis-à-vis *Non Replicated sampling designs*

Replicated sampling in each stratum has the advantages already mentioned in Volume I, as, for instance, allowing sample rotation. The area sample in Brazil was selected by substrata, but independent replicates were not selected.

Random Replicated vis-à-vis *Systematic Replicated Selection of Segments*

For each land use stratum, the *number* of *substrata* is equal to the number of segments of each replicate. And the number of sample segments in each substratum is equal to the number of replicates in the stratum.

In replicated sampling designs, the sample selection within each stratum can be done by selecting a random sample in each zone with or without replacement (*random replicated sampling*), or systematically by selecting replicates with identical selection intervals equal to the size of a substratum (*systematic replicated sampling*). These two selection methods are described in Chapter 7 of Volume I.

Although random replicated sampling within each *substratum (zone)* is recommended, systematic replicated sampling is often used because of its simplicity and because it also ensures a geographical spread of the sample of geographically ordered PSUs. Systematic sampling can cause a serious bias if the sampling interval should happen to coincide with a certain periodicity in the arrangement of the segments, but such periodicity is rare in area frame segment listings.

Estimation of Totals based on Substrata vis-à-vis *based on Replicates*

Estimates of a variable total can be obtained by using the replicates or by using the substrata. Unlike the estimation formulae using substrata, the formulae using replicates do not account for the substratification involved in the sample designs.

<u>*Characteristics of the Selection and Estimation Methods*</u>

- *Estimation by Substrata*. Replicated selection method. EPSEM selection of segments, without replacement, by substrata. Selection of one segment per replicate in each substratum. Two-stage sampling designs, and one-stage estimation formulae. Applied in the United States, Morocco and Pakistan.

- *Estimation by Substrata*. Systematic selection of segments by substrata, with no replicates. Successive difference variance estimation method. Single-stage sampling design and estimation formulae. Applied in Brazil.

- *Estimation by Strata*. Non-replicated design. Selection of PSU with PPS in each stratum and random selection of a segment within half of the PSU. Two-stage sampling design, and one-stage estimation formula. Applied in Albania.

- *Estimation by Strata*. Non-replicated design. Selection of PSU with PPS in each stratum and random selection of a segment within the PSU. Two-stage sampling design, and one-stage estimation formula. Applied in Colombia.

- *Estimation by Replicates*. Replicated selection method. Systematic selection of each replicate. EPSEM selection of one segment by substrata. Two-stage sampling designs, and one-stage estimation formulae. The estimation formulae used, for simplification purposes, do not account for the systematic selection procedures used. Applied in Honduras, Argentina, Costa Rica, the Dominican Republic, El Salvador and Guatemala.

3. DIRECT EXPANSION AREA SAMPLE ESTIMATORS

For the agricultural area sample surveys considered, each survey variable is defined in the *reporting units* (holdings or tracts), and the *sampling units* are the segments. Therefore, in order to define the estimator of a given *variable* for the survey area of interest it is necessary to define the variable in each sampling unit as a function of its values in a group of associated reporting units. In other words, for a given survey variable (defined in the set of reporting units), such definition (methods of association) refer to the way the data of the reporting units (holdings or tracts) are aggregated to define the segment totals.

As mentioned in Volume I, three methods are used to associate reporting units (holdings or tracts) to the sampling units (segments) to define the value of the survey variables for each segment: the *closed segment*, *weighted segment* and *open segment* methods.

Therefore, three general types of area sample estimators are used, namely *closed segment estimators, weighted segment estimators* and *open segment estimators* that correspond to the various methods that establish rules of association between the segments and the reporting units.

The type of estimator chosen for a survey variable depends on its reporting units (the way data pertaining to that variable have been collected).

Agricultural Area Sample Surveys:
Methods used to define survey variables in each segment

Sample unit	*Reporting unit*	*Method of association*
Segment	Holding	Weighted segment estimators
Segment	Holding	Open segment estimators
Segment	Tract	Closed segment estimators

The weighted and open segment estimators can be used for variables for which the *holding* is the reporting unit, and the closed estimator for variables for which the *tract* is the reporting unit.

For the *closed segment method*, the value of a variable in a segment is simply the sum of its values in each of the tracts of the segment.

For the *weighted segment method*, the variable in each *tract* is defined as the value of the variable in the total holding multiplied by a factor equal to the ratio between the area of the tract divided by the area of the holding. Then, the value of a variable in a *segment* is the sum of the variable in each of its tracts.

For the *open segment method*, the value of a variable in a *segment* is the sum of the variable in each of the holdings with headquarters included in the segment. As described in Volume I, the open segment method (or holding headquarters method) associates a segment to all holdings with headquarters included in the segment. For this purpose, rules have to be established to define a unique reference point for each holding, called the *headquarters*. The most common procedure is to define the headquarters as the dwelling (residence) of the holder.

Consider the following general notation for *survey designs with substrata*:

L = number of land use strata in the survey area.

h = stratum subindex ($h = 1, 2, \ldots, L$).

i = substratum subindex ($i = 1, 2, \ldots, L_h$).

j = segment subindex. Indicates a sequential number of a population segment or a sequential number of a sample segment.

k = agricultural holding or tract subindex.

N = total number of segments in the survey area.

N_h = number of segments in the h^{th} stratum.

N_{hi} = number of segments in the i^{th} substratum of the h^{th} stratum.

$$N = \sum_{h=1}^{L} N_h = \sum_{h=1}^{L} \sum_{i=1}^{L_h} N_{hi}$$

n = number of segments in the sample.

n_h = number of sample segments in the h^{th} stratum.

n_{hi} = number of sample segments in the i^{th} substratum of the h^{th} stratum (or number of replicates for estimators based in substrata).

$$n = \sum_{h=1}^{L} n_h = \sum_{h=1}^{L} \sum_{i=1}^{L_h} n_{hi}$$

$L_h = n_h / n_{hi} = N_h / N_{hi}$ = number of substrata in the h^{th} stratum.

$f_{hi} = n_{hi} / N_{hi}$ = sampling fraction for each segment of the i^{th} substratum in the h^{th} stratum.

$e_{hi} = N_{hi} / n_{hi}$ = expansion factor for each segment of the i^{th} substratum in the h^{th} stratum.

T_{hij} = number of tracts included in the j^{th} segment of the i^{th} substratum in the h^{th} stratum.

And consider the following notation for *stratified replicated survey designs*:

$n_{h\mu}$ = total number of sample segments in the μ^{th} replicate of the h^{th} stratum.

r_h = total number of replicates in the h^{th} stratum.

μ = replicate subindex ($\mu = 1,2,\ldots.. r_h$).

$n_h = r_h \cdot n_{h\mu}$

$f_h = n_h / N_h$ = sampling fraction in the h^{th} stratum.

$e_h = N_h / n_h$ = expansion factor in the h^{th} stratum.

$T_{h\mu j}$ = number of tracts included in the j^{th} segment of the μ^{th} replicate in the h^{th} stratum

Closed Segment Method for Survey Designs with Substrata

Assume that the variable X is defined in the set of all tracts and let:

X^c = denotes the variable, associated with X, but defined in the set of all segments, using the *closed segment method*.

X_{hijk} = value of the variable for the k^{th} tract, in the j^{th} segment of the i^{th} substratum in the hth stratum.

X^c_{hij} = value of the variable for the j^{th} segment of the i^{th} substratum in the h^{th} stratum (*closed segment method*).

$$X^c_{hij} = \sum_{k=1}^{T_{hij}} X_{hijk}$$

x_{hijk} = value of the variable for the k^{th} tract, in the j^{th} *sample* segment of the i^{th} substratum in the h^{th} stratum.

x^c_{hij} = value of the variable for the j^{th} *sample* segment of the i^{th} substratum in the h^{th} stratum (*closed segment method*).

$$x^c_{hij} = \sum_{k=1}^{T_{hij}} x_{hijk}$$

Weighted and Open Segment Methods for Survey Designs with Substrata

Assume now that the variable *X* is defined in the set of all holdings, and let:

X^w = denotes the variable, associated with *X*, but defined in the set of all segments, using the *weighted segment method*.

X_{hijk} = value of the variable corresponding to the **total** of the k^{th} *holding* partially or totally included in the j^{th} segment of the i^{th} substratum in the h^{th} stratum.

W_{hijk} = weight of the k^{th} tract in the j^{th} segment, of the i^{th} substratum in the h^{th} stratum (usually area of tract k divided by the total area of the holding corresponding to tract k).

X^w_{hij} = value of the variable for the j^{th} segment of the i^{th} substratum in the h^{th} stratum (*weighted segment method*).

$$X^w_{hij} = \sum_{k=1}^{T_{hij}} W_{hijk} \cdot X_{hijk} \qquad \text{if } T_{hij} > 0$$

$$X^w_{hij} = 0 \qquad \text{if } T_{hij} = 0$$

x_{hijk} = value of the variable corresponding to the **total** of the k^{th} *holding* partially or totally included in the j^{th} *sample* segment of the i^{th} substratum in the h^{th} stratum.

x^w_{hij} = value of the variable for the j^{th} *sample* segment of the i^{th} substratum in the h^{th} stratum (*weighted segment method*).

$$x^w_{hij} = \sum_{k=1}^{T_{hij}} W_{hijk} \cdot x_{hijk} \qquad \text{if } T_{hij} > 0$$

$x^{w}_{hij} = 0 \qquad \text{if } T_{hij} = 0$

$X^{o} =$ denotes the variable, associated with X, but defined in the set of all segments, using the *open segment method.*

$X^{o}_{hij} =$ value of the variable for the j^{th} segment of the i^{th} substratum in the h^{th} stratum (*open segment method*).

$\delta_{hijk} = 1,$ if the k^{th} holding headquarters is within the j^{th} segment of the i^{th} substratum in the h^{th} stratum.

$\delta_{hijk} = 0,$ otherwise.

$$X^{o}_{hij} = \sum_{k=1}^{T_{hij}} \delta_{hijk} \cdot X_{hijk} \qquad \text{if} \quad T_{hij} > 0$$

$$X^{o}_{hij} = 0 \qquad \text{if} \quad T_{hij} = 0$$

$x_{hijk} =$ value of the variable corresponding to the **total** of the k^{th} *holding* partially or totally included in the j^{th} *sample* segment of the i^{th} substratum in the h^{th} stratum.

$x^{o}_{hij} =$ value of the variable for the j^{th} *sample* segment of the i^{th} substratum in the h^{th} stratum (*open segment method*).

$$x^{o}_{hij} = \sum_{k=1}^{T_{hij}} \delta_{hijk} \cdot x_{hijk} \qquad \text{if} \quad T_{hij} > 0$$

$$x^{o}_{hij} = 0 \qquad \text{if} \quad T_{hij} = 0$$

Closed Segment Method for Stratified Replicated Designs

Assume that the variable X is defined in the set of all tracts and let:

$X^{c} =$ denotes the variable, associated with X, but defined in the set of all segments, using the *closed segment method.*

$X_{h\mu jk} =$ value of the variable for the k^{th} tract, in the j^{th} segment of the μ^{th} replicate in the h^{th} stratum.

$X^{c}_{h\mu j}$ = value of the variable for the $j^{\underline{th}}$ segment of the $\mu^{\underline{th}}$ replicate in the $h^{\underline{th}}$ stratum (*closed segment method*).

$$X^{c}_{h\mu j} = \sum_{k=1}^{T_{h\mu j}} X_{h\mu jk}$$

$x_{h\mu jk}$ = value of the variable for the $k^{\underline{th}}$ tract, in the $j^{\underline{th}}$ sample segment of the $\mu^{\underline{th}}$ replicate in the $h^{\underline{th}}$stratum.

$x^{c}_{h\mu j}$ = value of the variable for the $j^{\underline{th}}$ *sample* segment of the $\mu^{\underline{th}}$ replicate in the $h^{\underline{th}}$ stratum (*closed segment method*).

$$x^{c}_{h\mu j} = \sum_{k=1}^{T_{h\mu j}} x_{h\mu jk}$$

Weighted and Open Segment Methods for Stratified Replicated Survey Designs

Assume now that the variable X is defined in the set of all holdings and let:

X^{w} = denotes the variable, associated with X, but defined in the set of all segments, using the *weighted segment method*.

$X_{h\mu jk}$ = value of the variable corresponding to the **total** of the $k^{\underline{th}}$ *holding* partially or totally included in the $j^{\underline{th}}$ segment of the $\mu^{\underline{th}}$ replicate in the $h^{\underline{th}}$ stratum.

$x_{h\mu jk}$ = value of the variable corresponding to the **total** of the $k^{\underline{th}}$ *holding* partially or totally included in the $j^{\underline{th}}$ *sample* segment of the $\mu^{\underline{th}}$ replicate in the $h^{\underline{th}}$ stratum.

$W_{h\mu jk}$ = weight of the $k^{\underline{th}}$ tract in the $j^{\underline{th}}$ segment, of the $\mu^{\underline{th}}$ replicate in the $h^{\underline{th}}$ stratum (usually area of tract k divided by the total area of the holding corresponding to tract k).

$X^{w}_{h\mu j}$ = value of the variable for the $j^{\underline{th}}$ segment of the $\mu^{\underline{th}}$ replicate in the $h^{\underline{th}}$ stratum (*weighted segment method*).

$$X^{w}_{h\mu j} = \sum_{k=1}^{T_{h\mu j}} W_{h\mu jk} \cdot X_{h\mu jk} \qquad \text{if} \qquad T_{h\mu j} > 0$$

$$X^{w}_{h\mu j} = 0 \qquad \text{if} \qquad T_{h\mu j} = 0$$

$x^{w}_{h\mu j}$ = value of the variable for the j^{th} *sample* segment of the μ^{th} replicate in the h^{th} stratum (*weighted segment method*).

$$x^{w}_{h\mu j} = \sum_{k=1}^{T_{h\mu j}} W_{h\mu jk} \cdot x_{h\mu jk} \qquad \text{if} \qquad T_{h\mu j} > 0$$

$$x^{w}_{h\mu j} = 0 \qquad \text{if} \qquad T_{h\mu j} = 0$$

X^{o} = denotes the variable, associated with X, but defined in the set of all segments, using the *open segment method.*

$\delta^{*}_{h\mu jk} = 1$ if the k^{th} holding headquarters is within the j^{th} segment of the μ^{th} replicate in the h^{th} stratum.

$\delta^{*}_{h\mu jk} = 0$, otherwise.

$X^{o}_{h\mu j}$ = value of the variable corresponding to the j^{th} segment of the μ^{th} replicate in the h^{th} stratum (*open segment estimator*).

$$X^{o}_{h\mu j} = \sum_{k=1}^{T_{h\mu j}} \delta^{*}_{h\mu jk} \cdot X_{h\mu jk} \qquad \text{if} \qquad T_{h\mu j} > 0$$

$$X^{o}_{h\mu j} = 0 \qquad \text{if} \qquad T_{h\mu j} = 0$$

$x^{o}_{h\mu j}$ = value of the variable corresponding to the j^{th} segment of the μ^{th} replicate in the h^{th} stratum (*open segment method*).

$$x^{o}_{h\mu j} = \sum_{k=1}^{T_{h\mu j}} \delta^{*}_{h\mu jk} \cdot x_{h\mu jk} \qquad \text{if} \qquad T_{h\mu j} > 0$$

$$x^{o}_{h\mu j} = 0 \qquad \text{if} \qquad T_{h\mu j} = 0$$

A *direct expansion area sample estimator* of a given survey variable denotes the estimator that expands the data of each segment by the inverse of the sampling fraction in each stratum.

The following sections present the formulae for the closed, weighted and open segment direct expansion area sample estimators based on substrata or replicates.

4. DIRECT EXPANSION AREA SAMPLE ESTIMATORS FOR TOTALS BASED ON SUBSTRATA

Closed Segment Estimator based on Substrata

Assume that the variable X is be defined in the set of all tracts, and that X^c denotes the variable, associated with X, but defined in the set of all segments, using the *closed segment method*, and let:

X^c_{hi} = variable total for the i^{th} substratum in the h^{th} stratum.

$$X^c_{hi} = \sum_{j=1}^{N_{hi}} X^c_{hij} = N_{hi} \cdot \overline{X}^c_{hi}$$

X^c_h = variable total for the h^{th} stratum.

$$X^c_h = \sum_{i=1}^{L_h} X^c_{hi}$$

X^c = variable total for the survey area, based on the *closed segment method.*

$$X^c = \sum_{h=1}^{L} X^c_h$$

$$x^{c*}_{hij} = e_{hi} \cdot x^c_{hij}$$

$$\overline{x}^c_{hi} = (1/n_{hi}) \sum_{j=1}^{n_{hi}} x^c_{hij}$$

The estimator of the total X^c_{hi} is

$$\hat{X}^c_{hi} = N_{hi} \cdot \overline{x}^c_{hi} = e_{hi} \cdot x^c_{hi}$$

and the closed segment estimator of the total X^c, based on substrata, is

$$\hat{X}^c = \sum_{h=1}^{L} \hat{X}^c_h = \sum_{h=1}^{L} \hat{X}^c_{hi} = \sum_{h=1}^{L} \sum_{i=1}^{L_h} e_{hi} \cdot \sum_{j=1}^{n_{hi}} x^c_{hij} = \sum_{h=1}^{L} \sum_{i=1}^{L_h} \sum_{j=1}^{n_{hi}} x^{c*}_{hij}$$

Weighted Segment Estimator based on Substrata

Assume now that the variable X is defined in the set of all holdings, and that X^w denotes the variable, associated with X, but defined in the set of all segments, using the *weighted segment method* and let:

X^w_{hi} = variable total for the i^{th} substratum in the h^{th} stratum.

$$X^w_{hi} = \sum_{j=1}^{N_{hi}} X^c_{hij} = N_{hi} \cdot \overline{X}^w_{hi}$$

X^w_h = variable total for the h^{th} stratum.

$$X^w_h = \sum_{i=1}^{L_h} X^w_{hi}$$

X^w = variable total for the survey area, based on the *weighted segment method.*

$$X^w = \sum_{h=1}^{L} X^w_h$$

$$x^{w*}_{hij} = e_{hi} \cdot x^w_{hij}$$

$$\overline{x}^w_{hi} = (1/n_{hi}) \sum_{j=1}^{n_{hi}} x^w_{hij}$$

The estimator of the total X^w_{hi} is

$$\hat{X}^w_{hi} = N_{hi} \cdot \overline{x}^w_{hi} = e_{hi} \cdot x^w_{hi}$$

And the weighted segment estimator of the total X^w based on substrata, is

$$\hat{X}^w = \sum_{h=1}^{L} \hat{X}^w_h = \sum_{h=1}^{L} \sum_{i=1}^{L_h} \hat{X}^w_{hi} = \sum_{h=1}^{L} \sum_{i=1}^{L_h} e_{hi} \cdot \sum_{j=1}^{n_{hi}} x^w_{hij} = \sum_{h=1}^{L} \sum_{i=1}^{L_h} \sum_{j=1}^{n_{hi}} x^{w*}_{hij}$$

Open Segment Estimator based on Substrata

Assume now that the variable X is defined in the set of all holdings, and that X^o denotes the variable, associated with X, but defined in the set of all segments, using the *open segment method* and let:

X^o_{hi} = variable total for the i^{th} substratum in the h^{th} stratum.

$$X^o_h = \sum_{j=1}^{N_{hi}} X^o_{hij} = N_{hi} \cdot \bar{X}^o_{hi}$$

X^o_h = variable total for the h^{th} stratum.

$$X^o_h = \sum_{i=1}^{L_h} X^o_{hi}$$

X^o = variable total for the survey area, based on the *open segment method.*

$$X^o = \sum_{h=1}^{L} X^o_h$$

$$x^{o*}_{hij} = e_{hi} \cdot x^o_{hij}$$

$$\bar{x}^o_{hi} = (1/n_{hi}) \sum_{j=1}^{n_{hi}} x^o_{hij}$$

The estimator of the total X^o_{hi} is

$$\hat{X}^o_{hi} = N_{hi} \cdot \bar{x}^o_{hi} = e_{hi} \cdot x^o_{hi}$$

and the weighted segment estimator of the total X^o based on substrata, is

$$\hat{X}^o = \sum_{h=1}^{L} \hat{X}^o_h = \sum_{h=1}^{L} \sum_{i=1}^{L_h} \hat{X}^o_{hi} = \sum_{h=1}^{L} \sum_{i=1}^{L_h} e_{hi} \cdot \sum_{j=1}^{n_{hi}} x^o_{hij} = \sum_{h=1}^{L} \sum_{i=1}^{L_h} \sum_{j=1}^{n_{hi}} x^{o*}_{hij}$$

5. VARIANCE OF DIRECT EXPANSION AREA SAMPLE ESTIMATORS FOR TOTALS BASED ON SUBSTRATA

5.1 EPSEM Selection of Segments, Without Replacement, by Substrata

The three area sample estimators for totals based on substrata, defined in section 4, can be written as follows:

$$\hat{X} = \sum_{h=1}^{L} \sum_{i=1}^{L_h} \sum_{j=1}^{n_{hi}} x^{*}_{hij}$$

where:

$X = X^c,\ x^{*}_{hij} = x^{c*}_{hij} = e_{hi} \cdot x^{c}_{hij}$ for the closed segment estimator,

$X = X^w,\ x^{*}_{hij} = x^{w*}_{hij} = e_{hi} \cdot x^{w}_{hij}$ for the weighted segment estimator, and

$X = X^o,\ x^{*}_{hij} = x^{o*}_{hij} = e_{hi} \cdot x^{o}_{hij}$ for the open segment estimator.

Then, the estimated variance formulae for the three direct area sample estimators based on substrata can be written:

$$\hat{V}(\hat{X}) = \sum_{h=1}^{L} \sum_{i=1}^{L_h} \hat{V}(\hat{X}_{hi}) =$$

$$= \sum_{h=1}^{L} \sum_{i=1}^{L_h} (1 - f_{hi}) \cdot (N^2_{hi} / n_{hi}) \cdot [1/(n_{hi} - 1)] \cdot \sum_{j=1}^{n_{hi}} (x_{hij} - \bar{x}_{hi})^2 =$$

$$= \sum_{h=1}^{L} \sum_{l=1}^{L_h} (1 - f_{hi}) \cdot [n_{hi} / (n_{hi} - 1)] \cdot \left[(\sum_{j=1}^{n_{hi}} x^{*2}_{hij}) - (1/n_{hi}) \cdot (\sum_{j=1}^{n_{hi}} x^{*}_{hij})^2 \right]$$

Observe that the formula given in Volume I, section 4.3, does not include the finite population correction factor $(1 - f_{hj})$.

5.2 Systematic Selection of Segments by Substrata, Without Replicates. Variance Formulae Using the Successive Difference Method

The variance formulae using the successive difference model for the three direct area sample estimators based on substrata (cf. section 4) are as follows (Kish, 1965, Section 4.1.2, p. 119):

$$\hat{V}(\hat{X}) = \sum_{h=1}^{L} \sum_{i=1}^{L_h} N^2_{hi}\ \hat{V}(\bar{x}_{hi}) =$$

$$= \sum_{h=1}^{L} \sum_{i=1}^{L_h} N_{hi}^2 \cdot [(1-f_{hi})/2n_{hi} \cdot (n_{hi}-1)] \sum_{j=1}^{n_{hi}-1} (x_{hij} - x_{hi(j+1)})^2$$

where

$X = X^c$ and $x_{hij} = x_{hij}^c$ for the closed segment estimator,

$X = X^w$ and $x_{hij} = x_{hij}^w$ for the weighted segment estimator, and

$X = X^o$ and $x_{hij} = x_{hij}^o$ for the open segment estimator.

6. DIRECT EXPANSION AREA SAMPLE ESTIMATORS FOR TOTALS BASED ON REPLICATES BY STRATA

Direct Expansion Area Sample Estimators based on Replicates

Assume that the variable Y_h is defined in all segments of the h^{th} stratum and that $Y = \sum_{h=1}^{L} Y_h$ is defined in all segments of the total survey area.

Then, an unbiased estimate of the variable total Y_h in the stratum (called the estimate based on the replicates) is the average of the r_h estimates of Y_h, each obtained by considering only one replicate. In fact, let:

Y = variable total for the total survey area.

Y_h = variable total in the h^{th} stratum.

$Y_{h\mu j}$ = value of the variable Y for the j^{th} segment of the μ^{th} replicate in the h^{th} stratum.

$y_{h\mu j}$ = value of the variable for the j^{th} sample segment of the μ^{th} replicate in the h^{th} stratum.

$$y^*_{h\mu j} = r_h \cdot e_h \cdot y_{h\mu j}$$

$$y^*_{h\mu} = \sum_{j=1}^{n_{h\mu}} y^*_{h\mu j} = \sum_{j=1}^{n_{h\mu}} r_h \cdot e_h \cdot y_{h\mu j}$$

The estimator of Y_h based only on the μ^{th} replicate is the following:

$$\hat{Y}_{h,\mu} = N_h \cdot (1/n_{h\mu}) \cdot \sum_{j=1}^{n_{h\mu}} y_{h\mu j} = y^*_{h\mu}$$

Then, an unbiased estimator of Y_h that combines the estimators derived from each of the r_h replicates is the following (cf. Kish, 1965, formula 4.4.1):

$$\hat{Y}_h = (1/r_h) \cdot \sum_{\mu=1}^{r_h} \hat{Y}_{h,\mu} = (1/r_h) \cdot \sum_{\mu=1}^{r_h} y^*_{h\mu} =$$

$$= (1/r_h) \cdot \sum_{\mu=1}^{r_h} \sum_{j=1}^{n_{h\mu}} y^*_{h\mu j} = \sum_{\mu=1}^{r_h} \sum_{j=1}^{n_{h\mu}} e_h \cdot y_{h\mu j}$$

The estimator of Y based on replicates is

$$\hat{Y} = \sum_{h=1}^{L} \hat{Y}_h = \sum_{h=1}^{L} (1/r_h) \cdot \sum_{\mu=1}^{r_h} \hat{Y}_{h,\mu} = \sum_{h=1}^{L} (1/r_h) \cdot \sum_{\mu=1}^{r_h} y^*_{h\mu} =$$

$$= \sum_{h=1}^{L} (1/r_h) \cdot \sum_{\mu=1}^{r_h} \sum_{j=1}^{n_{h\mu}} y^*_{h\mu j} = \sum_{h=1}^{L} \sum_{\mu=1}^{r_h} \sum_{j=1}^{n_{h\mu}} e_h \cdot y_{h\mu j}$$

Closed Segment Estimator based on Replicates

Assume that the variable X is defined in the set of all tracts, X^c is the variable total over all strata, and X^c_h is the variable total for the h^{th} stratum, as already defined. Then, for $Y = X^c$, $y = x^c$ and $y^* = x^{c*}$ in the general formula above, the estimator of X^c_h based only on the μ^{th} replicate is $x^{c*}_{h\mu}$, and an unbiased estimator of X^c_h that combines the estimators derived from each of the r_h replicates is the following:

$$\hat{X}^c_h = (1/r_h) \cdot \sum_{\mu=1}^{r_h} \hat{X}^c_{h,\mu} = (1/r_h) \cdot \sum_{\mu=1}^{r_h} x^{c*}_{h\mu} =$$

$$= (1/r_h) \cdot \sum_{\mu=1}^{r_h} \sum_{j=1}^{n_{h\mu}} x^{c*}_{h\mu j} = \sum_{\mu=1}^{r_h} \sum_{j=1}^{n_{h\mu}} e_h \cdot x^c_{h\mu j}$$

And the estimator of X^c is

$$\hat{X}^c = \sum_{h=1}^{L} \hat{X}^c_h = \sum_{h=1}^{L} (1/r_h) \cdot \sum_{\mu=1}^{r_h} \hat{X}^c_{h,\mu} = \sum_{h=1}^{L} (1/r_h) \cdot \sum_{\mu=1}^{r_h} x^{c*}_{h\mu} =$$

$$= \sum_{h=1}^{L} (1/r_h) \cdot \sum_{\mu=1}^{r_h} \sum_{j=1}^{n_{h\mu}} x_{h\mu j}^{c*} = \sum_{h=1}^{L} \sum_{\mu=1}^{r_h} \sum_{j=1}^{n_{h\mu}} e_h \cdot x_{h\mu j}^{c}$$

Weighted Segment Estimator based on Replicates

Assume now that the variable *X* is defined in the set of all holdings, X^w is the variable total over all strata, and X_h^W is the variable total for the h^{th} stratum. Then, for $Y = X^W$, $y = x^W$ and $y = x^{W*}$ in the general formula, the estimator of X_h^W based only on the μ^{th} replicate is $x_{h\mu}^{W*}$, and an unbiased estimator of X_h^W that combines the estimators derived from each of the r_h replicates is the following:

$$\hat{X}_h^W = (1/r_h) \cdot \sum_{\mu=1}^{r_h} \hat{X}_{h,\mu}^W = (1/r_h) \cdot \sum_{\mu=1}^{r_h} x_{h\mu}^{W*} =$$

$$= (1/r_h) \cdot \sum_{\mu=1}^{r_h} \sum_{j=1}^{n_{h\mu}} x_{h\mu j}^{W*} = \sum_{\mu=1}^{r_h} \sum_{j=1}^{n_{h\mu}} e_h \cdot x_{h\mu}^{W}$$

And the estimator of X^w is

$$\hat{X}^W = \sum_{h=1}^{L} \hat{X}_h^W = \sum_{h=1}^{L} (1/r_h) \cdot \sum_{\mu=1}^{r_h} \hat{X}_{h,\mu}^W = \sum_{h=1}^{L} (1/r_h) \cdot \sum_{\mu=1}^{r_h} x_{h\mu}^{W*} =$$

$$= \sum_{h=1}^{L} (1/r_h) \cdot \sum_{\mu=1}^{r_h} \sum_{j=1}^{n_{h\mu}} x_{h\mu j}^{W*} = \sum_{h=1}^{L} \sum_{\mu=1}^{r_h} \sum_{j=1}^{n_{h\mu}} e_h \cdot x_{h\mu j}^{W}$$

Open Segment Estimator based on Replicates

Assume that the variable *X* is defined in the set of all holdings, X^o is the variable total over all strata, and X_h^o is the variable total for the h^{th} stratum, as already defined. Then, for $Y = X^o$, $y = x^o$ and $y^* = x^{o*}$ in the general formula, the estimator of X_h^o based only on the μ^{th} replicate is $x_{h\mu}^{o*}$, and an unbiased estimator of X_h^o that combines the estimators derived from each of the r_h replicates is the following:

$$\hat{X}_h^o = (1/r_h) \cdot \sum_{\mu=1}^{r_h} \hat{X}_{h,\mu}^o = (1/r_h) \cdot \sum_{\mu=1}^{r_h} x_{h\mu}^{o*} =$$

$$= (1/r_h) \cdot \sum_{\mu=1}^{r_h} \sum_{j=1}^{n_{h\mu}} x^{o*}_{h\mu j} = \sum_{\mu=1}^{r_h} \sum_{j=1}^{n_{h\mu}} e_h \cdot x^{o}_{h\mu}$$

And the estimator of X^o is

$$\hat{X}^o = \sum_{h=1}^{L} \hat{X}^o_h = \sum_{h=1}^{L} (1/r_h) \cdot \sum_{\mu=1}^{r_h} \hat{X}^o_{h,\mu} = \sum_{h=1}^{L} (1/r_h) \cdot \sum_{\mu=1}^{r_h} x^{o*}_{h\mu} =$$

$$= \sum_{h=1}^{L} (1/r_h) \cdot \sum_{\mu=1}^{r_h} \sum_{j=1}^{n_{h\mu}} x^{o*}_{h\mu j} = \sum_{h=1}^{L} \sum_{\mu=1}^{r_h} \sum_{j=1}^{n_{h\mu}} e_h \cdot x^{o}_{h\mu j}$$

7. VARIANCE OF DIRECT EXPANSION AREA SAMPLE ESTIMATORS FOR TOTALS BASED ON REPLICATES

The variance formulae for the three direct area sample estimators based on replicates assumes that each replicate is selected with the same method and sampling fraction. The formulae also apply in the case in which there is a systematic selection for each replicate (random group method).

As already mentioned, the three area sample estimators based on replicates, defined in section 6, can be written as follows:

$$\hat{Y} = \sum_{h=1}^{L} \hat{Y}_h = \sum_{h=1}^{L} (1/r_h) \cdot \sum_{\mu=1}^{r_h} \hat{Y}_{h,\mu} = \sum_{h=1}^{L} (1/r_h) \cdot \sum_{\mu=1}^{r_h} y^{*}_{h\mu} =$$

$$= \sum_{h=1}^{L} (1/r_h) \cdot \sum_{\mu=1}^{r_h} \sum_{j=1}^{n_{h\mu}} y^{*}_{h\mu j} = \sum_{h=1}^{L} \sum_{\mu=1}^{r_h} \sum_{j=1}^{n_{h\mu}} e_h \cdot y_{h\mu j}$$

where:

$y_{h\mu j}$ = value of the variable for the j^th^ *sample* segment of the μ^{th} replicate in the h^{th} stratum.

$$y^{*}_{h\mu j} = r_h \cdot e_h \cdot y_{h\mu j}$$

$$y^{*}_{h\mu} = \sum_{j=1}^{n_{h\mu}} y^{*}_{h\mu j}$$

$Y = X^c$, $y^*_{h\mu j} = x^{c*}_{h\mu j} = r_h \cdot e_h \cdot x^c_{h\mu j}$ for the *closed segment estimator,*

$Y = X^w$, $y^*_{h\mu j} = x^{w*}_{h\mu j} = r_h \cdot e_h \cdot x^w_{h\mu j}$ for the *weighted segment estimator,*

$Y = X^o$, $y^*_{h\mu j} = x^{o*}_{h\mu j} = r_h \cdot e_h \cdot x^o_{h\mu}$ for the *open segment estimator.*

Then, the estimated variance formulae for the three direct area sample estimators based on replicates can be written (cf. Kish, 1965, formula 4.4.2):

$$\hat{V}(\hat{Y}) = \sum_{h=1}^{L} \hat{V}(\hat{Y}_h) = \sum_{h=1}^{L} \hat{V}\left[(1/r_h) \cdot \sum_{\mu=1}^{r_h} y^*_{h\mu}\right] =$$

$$= \sum_{h=1}^{L} \left[(1-f_h)/r_h \cdot (r_h - 1)\right] \cdot \left[(\sum_{\mu=1}^{r_h} y^{*2}_{h\mu}) - (1/r_h) \cdot (\sum_{\mu=1}^{r_h} y^*_{h\mu})^2\right]$$

8. RATIO AREA SAMPLE ESTIMATORS AND THEIR VARIANCES

8.1 Within-Survey Ratios

This is an area sample ratio estimator which measures the proportion of one variable to another variable from the same survey round. The numerator and denominator totals are calculated at the stratum level and then summed over strata to obtain the province-level ratio. Examples are yields for crops, milk production per cow, litter rates for hogs, etc. Let:

$\hat{Y}$ and $\hat{X}$ = Direct expansion estimators (closed segment, weighted segment or open segment) of the variables Y and X, based on substrata or replicates.

$\hat{R}_w$ = Ratio estimator for the variables $\hat{Y}$ and $\hat{X}$, from the same survey round.

$$\hat{R}_w = \hat{Y} / \hat{X}$$

An approximation of the estimated variance of $\hat{R}$ is the following:

$$\hat{V}(\hat{R}_w) = \hat{R}^2_w \left[(\hat{V}(\hat{Y}) / \hat{Y}^2) + (\hat{V}(\hat{X}) / \hat{X}^2) - 2\, \hat{C}ov(Y,X) / \hat{Y} \cdot \hat{X})\right]$$

where

$\hat{V}(\hat{Y})$ and $\hat{V}(\hat{X})$ = Estimators of the variances of $\hat{Y}$ and $\hat{X}$, using the closed segment, weighted segment or open segment methods, and based on substrata or replicates.

$$\hat{C}ov(Y,X) = \sum_{h=1}^{L} \sum_{i=1}^{L_h} (1 - f_{hi}) \cdot [n_{hi} / (n_{hi} - 1)] \cdot \sum_{j=1}^{n_{hi}} (y^*_{hij} - \bar{y}^*_{hi}) \cdot (x^*_{hij} - \bar{x}^*_{hi}) =$$

= Covariance estimator for Y and X, using the closed segment, weighted segment or open segment methods, and *based on substrata.*

$$\hat{C}ov(Y,X) = \sum_{h=1}^{L} (1 - f_h) \cdot [r_h / (r_h - 1)] \cdot \sum_{\mu=1}^{r_h} (y^*_{h\mu} - \bar{y}^*_h) \cdot (x^*_{h\mu} - \bar{x}^*_h) =$$

= Covariance estimator for Y and X, using the closed segment, weighted segment or open segment methods, and *based on replicates.*

Observe that the formulae given in Volume I, section 4.4.1, do not include the finite population correction factor.

8.2 Between-Survey Ratios

This is an area sample ratio estimator which measures the percentage change in a characteristic from the previous (base) survey round. The current and previous year total estimates are calculated at the land use stratum level and then summed over strata to produce the province/state level ratio. Only matching area sample segments may be used in the calculation: for example, if 20% of replicates are rotated each year, the between-survey ratios are calculated using only 80% of the current year's segment data. Usual examples of between-survey ratios are percentage changes in totals (livestock inventory levels, crop planted areas, and number of holdings, etc.). Let:

$$\hat{R}_b = \hat{Y}_{current} / \hat{X}_{base}$$

where:

$\hat{R}_b$ = Between-survey ratio estimator of the variable $\hat{Y}$ in the *current year* and $\hat{X}$ in the previous (base) year (usually the characteristic is the same, but Y and X have different values).

n^{α}_{hi} = number of sample segments in the i^{th} substratum of the h^{th} stratum that belong both to the current year and base year survey rounds.

r^{α}_h = total number of replicates in the h^{th} stratum, common to both the current year and the base survey round.

$$\hat{Y}_{current} = \sum_{h=1}^{L} \sum_{i=1}^{L_h} (n_{hi} / n^{\alpha}_{hi}) \cdot \sum_{j=1}^{n^{\alpha}_{hi}} \sum_{k=1}^{T_{hij}} y^*_{hijk}$$

= Direct expansion estimator (closed segment, weighted segment or open segment) of the variable Y, based on *substrata*.

or

$$\hat{Y}_{current} = \sum_{h=1}^{L} \sum_{\mu=1}^{r_h^{\alpha}} (r_h / r_h^*) \cdot \sum_{j=1}^{n_{h\mu}} \sum_{k=1}^{T_{h\mu j}} y^*_{h\mu jk}$$

= Direct expansion estimator (closed segment, weighted segment or open segment) of the variable Y, based on *replicates*.

The definition of $\hat{X}_{base}$ is analogous to $\hat{Y}_{current}$.

The variance formulae are very similar to the variances for the within-survey ratios. The variances and the covariances are calculated substituting n_{hi} and r_h, by n^{α}_{hi} and r^{α}_h respectively, and considering also the factors $(n_{hi} / n^{\alpha}_{hi})$ or (r_h / r^{α}_h) accordingly.

9. DUAL FRAME ESTIMATORS

For survey variables with a distribution such that a large percentage of the total is concentrated in a relatively small number of the holdings (called *special holdings*), it is often convenient to use dual estimators in order to improve the precision of the direct area sample estimators. The complementary list frame of special holdings ensures the inclusion of those holdings which make a significant contribution to the total estimate of some important survey variables. As already mentioned, such a list of special holdings may consist, for instance, of those holdings with the largest total area, those with the largest area for a given crop, those with the largest number of livestock and poultry, those with the largest revenues, those with the largest number of agricultural workers, those corresponding to a localized production and those concentrating on highly specialized types of production. In addition, the technical difficulties of adding a short list of special agricultural holdings (the area frame component) to an area sample design are relatively minor (cf. Volume I).

Therefore, to obtain the estimates of a number of important survey variables, it is generally recommended to use a relatively short list of special holdings, denoted by Ω, to be completely enumerated, in combination with the area sample. This is the type of dual frame estimators described in this section. In fact, these are the most practical multiple frame methods for developing countries.

The estimator commonly used is called *screening multiple frame estimator* because the tracts in the area sample are "screened" and only those not included in the list frame are enumerated.

The screening dual frame estimator $\hat{X}$ of a given variable total *X* defined in the set of all holdings, combines the area sample estimator with the list frame data:

$$\hat{X} = X_{list} + \hat{X}_{area,non\ list}$$

where:

X_{list} : indicates the data collected from all holdings of Ω (the list of special holdings).

It should be noted that instead of the constant X_{list} the estimator $\hat{X}_{list}$ could be used. In this case, a stratified simple random sample of Ω is generally used (e.g. the dual frame design used in Brazil).

$\hat{X}_{area,\,non\ list}$: indicates the "weighted" estimator corresponding to the tracts in the area sampling frame, not included in Ω.

The estimator $\hat{X}_{area,\,non\ list}$, using the weighted segment estimator based on substrata, is defined by the following identity:

$$\hat{X}_{area,\,non\ list} = \sum_{h=1}^{L} \sum_{i=1}^{L_h} e_{hi} \cdot \sum_{j=1}^{n_{hi}} \sum_{k=1}^{T^*_{hij}} x_{hijk} =$$

$$= \sum_{h=1}^{L} \sum_{i=1}^{L_h} e_{hi} \cdot \sum_{j=1}^{n_{hi}} x^{w}_{hij} = \sum_{h=1}^{L} \sum_{i=1}^{L_h} \sum_{j=1}^{n_{hi}} x^{w*}_{hij}$$

where:

W^*_{hijk} = weight of the k^{th} tract in the j^{th} segment, of the i^{th} substratum of the h^{th} stratum (usually area of tract k divided by the total area of the holding corresponding to tract k).

W^*_{hijk} = 0 if the k^{th} tract in the j^{th} segment, of the i^{th} substratum of the h^{th} stratum belongs to a holding of Ω.

T^*_{hij} = number of tracts included in the j^{th} segment of the i^{th} substratum of the stratum h, that do not belong to Ω.

$$x^{w}_{hij} = \sum_{k=1}^{T^*_{hij}} W^*_{hijk} \cdot x_{hijk} \qquad \text{if} \qquad T^*_{hij} > 0$$

$x^{w}_{hij} \;=\; 0 \qquad\qquad \text{if} \qquad T^{*}_{hij} = 0$

$x^{w*}_{hij} = e_{hi} \cdot x^{w}_{hij}$

$\hat{V}(\hat{X}_{area,\, non\, list})$, the estimated variance of $\hat{X}_{area,non\, list}$ is equal to:

$$\sum_{h=1}^{L} \sum_{i=1}^{L_h} (1-f_{hi}) \cdot \left[n_{hi} / (n_{hi}-1)\right] \cdot \left[(\sum_{j=1}^{n_{hi}} x^{w*^2}_{hij}) - (1/n_{hi}) \cdot (\sum_{j=1}^{n_{hi}} x^{w*}_{hij})^2 \right]$$

Therefore, when the list of special holdings is completely enumerated, the variance of the screening dual frame estimator is equal to the variance of the area sample estimator considered, but the CV of the dual frame estimator will always be less than or equal to the CV of the area sample estimator. Therefore, dual frame surveys provide more precise estimates than area sample surveys for those variables partially accounted for in the list of special holdings.

PART III

SPECIAL PURPOSE (LAND USE, CROP PRODUCTION OR CROP AREA) AGRICULTURAL SURVEYS IN SOME EUROPEAN COUNTRIES BASED ON AREA SAMPLING METHODS

Surveys that do not involve interviews with farmers

Part III includes methodological descriptions of *special purpose* agricultural survey programmes in European countries which do not involve data collection through personal interviews with farmers and are based on area sampling methods using square segments. The TER-UTI land use survey in France, the crop production survey in Italy and the crop area survey conducted in some regions of Spain are described. The special purpose surveys described are based on area sampling methods different from those used for general purpose agricultural survey programmes, having more restricted objectives.

CHAPTER 12

FRANCE

LAND USE AGRICULTURAL SURVEY PROGRAMME BASED ON AREA FRAME SAMPLING METHODS (1970-1997)

1. INTRODUCTION

This chapter describes the land use survey conducted annually in France, called *TER-UTI*, based on point sampling methods. The TER-UTI survey is the basic source for the estimation of land areas in France. In addition, a succinct reference to the *Arable Land Production Survey*, based on a subsample of the TER-UTI sample, follows in section 5.

In France, almost all statistics on the agricultural sector are collected through National Agricultural Censuses about every nine years and annual sample surveys based on list sampling designs. The sampling unit of these surveys is the holding (farm) and data are collected through interviews with farmers. In particular, these survey programmes provide estimates of the farm area which covers about 64% of the national territory: almost all the utilized agricultural area (UAA), a large portion of fallow land, moors and heath land, but a smaller percentage of woodland and an even smaller percentage of non-agricultural area.

2. THE TER-UTI NATIONAL SURVEY

The TER-UTI survey is conducted annually by enumerators who visit more than 550 000 "points" (square *segments* of 9 m^2) in order to report, for each "point", one and only one of the 81 pre-established *physical characteristics* (e.g. area with oats), and also one and only one of the 25 pre-established *functional characteristics* (e.g. agricultural area). The list of possible *physical* and *functional* characteristics - the nomenclatures - is given at the end of this summary.

The TER-UTI area sample survey covers the whole French territory and allows the study of a variety of agricultural, forestry, environmental and urban problems. Survey results are published in September each year.

3. SURVEY OBJECTIVES

The objective of the survey, as defined by the Ministry of Agriculture, is to obtain estimates of land cover and to monitor the changes between agricultural and non-agricultural land.

The first experiments were carried out on cadastral plots of the Area Control survey from 1946 to 1969. Local committees in each town were given the task of organizing an exhaustive survey of all tracts, with the help of cadastral maps. Difficulties were encountered because of the poor quality of the cadastre and the difference between agricultural fields and cadastral plots. In areas where the cadastre was not updated, aerial photographs were used instead of cadastral maps: this was the beginning of the change from cadastral maps to aerial photographs, which eventually led to the TER-UTI survey.

The TER-UTI survey was adjusted during the 1970s, and from 1980 it was established as a regular survey programme.

The objectives of TER-UTI national land cover survey are the following:

- to establish a periodic system for monitoring the evolution of the agricultural and non-agricultural territory: crop rotation, changes and land cover;
- to obtain yearly data for each of the 96 *departments* (basic administrative divisions with an area of about 6 000 km^2) and each of the 22 *regions*;
- to provide an area sampling frame that could also be used for specific statistical studies (orchard surveys, arable land yield surveys, etc.).

4. THE SAMPLE DESIGN

A two-stage systematic probability sampling design with no stratification is used, as described in the following sections and illustrated in Figure 12.1.

The area sampling frame survey established in 1980 was renewed in 1991-1992.

The area sample design is basically that described by Fournier (1972).

4.1 Area Frame Construction and Sample Selection

***First sampling stage*: *Systematic Aligned Selection of PSUs*.**

For constructing the first stage sampling frame, a regular grid of square cells (each cell in the direction North-South and East-West) - of 10.8 km x 10.8 km - is overlaid on the national territory. A PSU is a square of 10.8 km x 10.8 km, and is called a *position*. Each cell is divided into 36 PSUs, which are equal squares except for cells partially included in the territory because of crossing boundaries.

The same eight fixed positions are considered for selection in each cell. In the area of *Paris and neighbourhood* and in the *territory of Belfort*, all of these eight positions (1 to 8 in Figure 12.1) are selected in the first sampling stage. In the rest of the French territory, only four positions (1 to 4 in Figure 12.1, at a distance of approximately 6 km) are selected for the first sampling stage.

Aerial photography for each selected position is examined and the photo with the centre nearest to the position is selected. A total of 15 389 aerial photos corresponding to the above-mentioned systematic selection of *positions* constitute the selected PSUs. Therefore a sampling fraction of approximately 1/10 is used in the first sampling stage.

Second sampling stage: *Systematic Aligned Selection of Square Segments*

For each selected PSU (position) there is a corresponding aerial photograph in scale 1:5 000 prepared by the National Geographic Institute. A PSU aerial photo, covering an area of 2 km x 2 km, is shown in Figure 12.1.

Each PSU can be considered as formed by 360 000 *square segments of* 9 m^2 (3 m x 3 m).

A square grid of 36 points (300 m apart) is overlaid on each aerial photograph, and the 36 segments (points) are systematically selected, as shown in Figure 12.1, where each point is represented by a cross.

Therefore, the second stage sampling fraction is around 1/10 000, and the total sampling fraction is approximately equal to 1/100 000.

The number of selected points (segments) in France is equal to 555 903 and the continental total area of France equals 549 192 km^2.

4.2 Survey Questionnaire

As already mentioned, to each sample *point* the enumerator assigns one physical occupational category as well as a functional use category concerning the socio-economic function of the segment, as shown at the end of this summary.

The *physical nomenclature* divides the land cover into five major groups:

- permanent water and humid zones;
- rock, pebbles and sand;
- wooded area;
- utilized agricultural area;
- other areas.

These groups are subdivided into a total of 81 categories.

The *functional nomenclature* has three major groups:

- primary production;
- secondary production;
- services and others.

These groups are subdivided into a total of 25 categories.

The TER-UTI survey studies two discrete variables, with 81 and 25 possible values respectively.

4.3 Field Data Collection

The survey involves the completion of questionnaires by enumerators during the field data collection. It does not involve farmer interviews.

The work of the enumerator consists in finding the location of the selected photograph on a map, locating the photograph on the ground, finding the exact position of each of the 36 points of the photograph and observing the segment to complete the questionnaire.

The segment consists of a small area of about 9 m^2 around the point. From 1993, the observation area has been extended to a circle whose radius is 15 m (700 m^2) in areas with heterogeneous land cover, such as transition areas, i.e areas that are part of the UAA (utilized agricultural area) and become part of the non-agricultural domain owing to urban and industrial growth and farmland abandonment. The point is classified as the cover or land use that occupies the majority of the observed area.

Points located on field borders can generate many ambiguous cases. The common method used to solve these problems is the random selection of one of the two adjacent fields.

For each sample segment, the *physical nomenclature code* is obtained by the enumerator through direct observation of the area. The determination of the *functional nomenclature* code is less direct and sometimes more difficult to establish since it refers to the socio-economic function of the segment. However, quite often the physical nomenclature is sufficient to determine the functional nomenclature code. When this is not possible, it is necessary for the enumerator to observe also the neighbouring area of the segment. This is the case, for example, of a field belonging to a research institute or a vocational school, that would be classified as "education and research".

Data collection materials used by enumerators

- topographic maps;
- 1:5 000 enlargements of aerial photographs with the 36-point grid overlaid;
- questionnaires;
- pencil;
- instruction leaflet.

Besides these basic instruments, the enumerators are given specific information sheets on the determination of Graminae (cereals and grass), on the leguminous plants found in grassland and on the criteria used to classify the different types of grassland.

Training of enumerators. The local statisticians are in charge of training the enumerators with the help of an “instructor package” which includes a set of 100 colour slides emphasizing difficult cases.

Data collection period and dates. The data collection is carried out between 1 June and 30 September, depending on the vegetative state of the crops. The enumerators observe the points on the ground and the photographs only once a year at a date when all the crops are growing, including summer crops such as maize, soybeans, sugar beet and sunflower.

Quality control. Two types of data quality control are carried out: consistency control and systematic control.

Consistency control refers to checking whether the land cover collected for a specific year is consistent with the land cover of the previous year. For example, it is very unlikely that a field of wheat will become a wooded area or a vineyard the next year. These cases are checked by a supervisor.

Systematic control is the check of four points randomly selected on one out of four of the selected photographs.

The main errors detected are:

- location errors (the enumerator locates the point in a wrong place);
- observation errors;
- interpretation errors (misunderstanding of the instructions regarding particular cases).

4.4 Estimation Methods and their Precision

The expansion factor v_h of a TER-UTI point in a department is given by the ratio of the area of the department, as measured by the National Geographic Institute for this purpose, and the number of points in the department.

The area Z_{hc} of land use category c in the h^{th} department is estimated as follows:

$$z_{hc} = v_h \, n_{hc},$$

where:

z_{hc} = estimated area of the category c;
n_{hc} = number of points of the category c;
v_h = expansion factor for a TER-UTI point.

Each sample point represents about 1 km^2 (the expansion factor), except in small departments where it represents 0.5 km^2.

The variance had always been estimated by interpreting the actual two-stage sampling design as a single-stage sample selection with replacement, which leads to the use of the binomial random variable:

$$V(z_{hc}) = z_{hc} \, (Z_h - z_{hc}) \, / \, n_h,$$

where:

Z_h = total area of the h^{th} department; and
n_h = number of points in the h^{th} department.

However, from 1997, a variance estimator proposed by Fournier (1972) that considers a two-stage systematic area sample design with a distance threshold, has been implemented and compared with the previous estimator (cf. Amorich, *et al.*, 1997). The formula interprets the actual design as having a first-stage EPSEM systematic selection of PSUs of equal size (3.24 km^2) - N = 169 503, n = 15 389; and a second stage EPSEM systematic selection of segments.

5. THE ARABLE LAND PRODUCTION SURVEY

The objective of the arable land production survey is to estimate the production of major crops at the department, region and country levels.

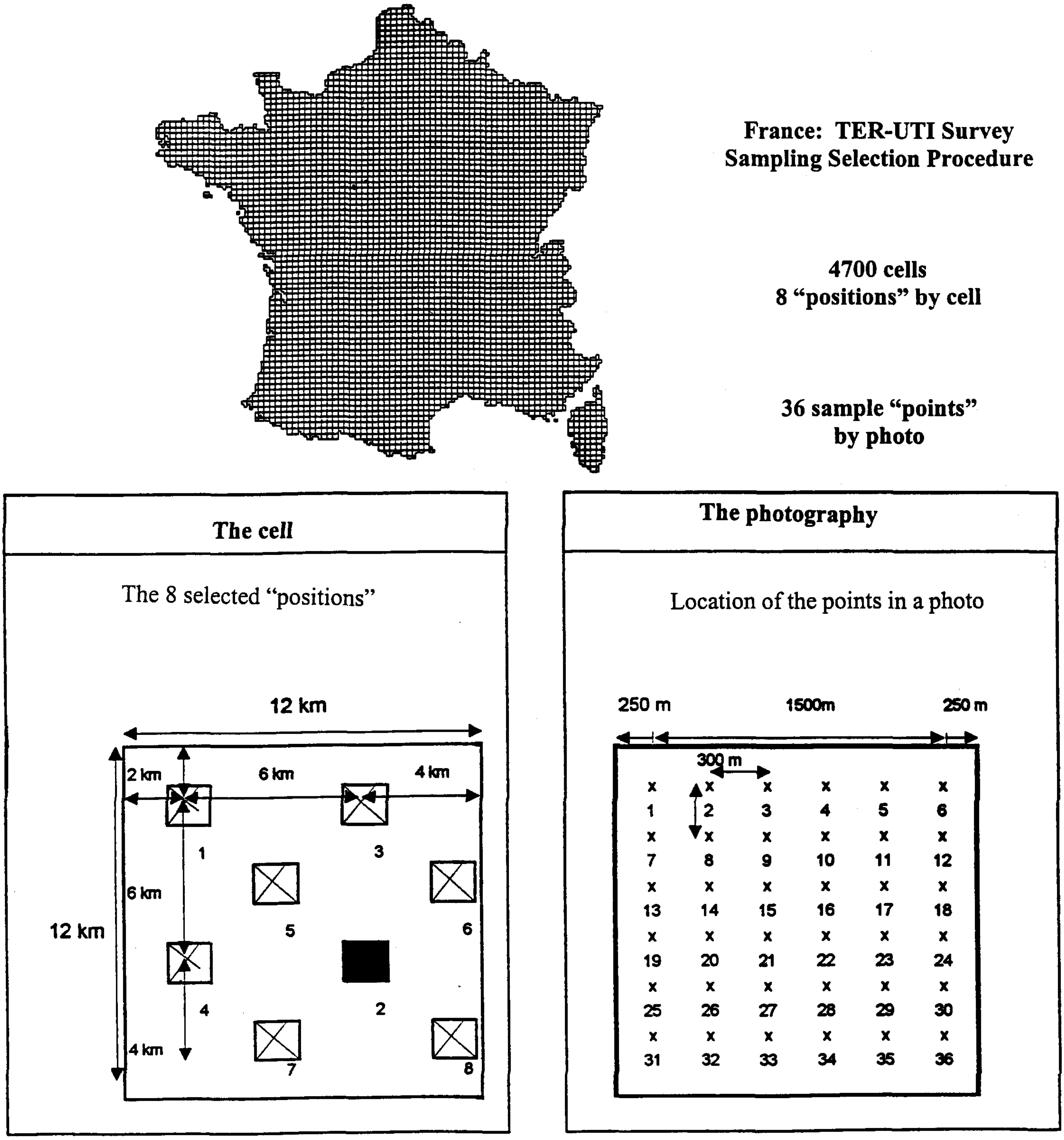
Production estimation involves two aspects: area estimation and crop yield estimation for main crops.

- Area estimations are based mainly on the TER-UTI.
- Yields are estimated through a specific survey carried out on a sample of agricultural holdings. A total of 300 points in each department is selected from the TER-UTI points falling on arable land. These points correspond to a maximum of 300 holdings (some points can fall on the same holding) which represent about 4% of holdings with arable land.

A questionnaire is sent to the farmers by the end of spring, in order to obtain the area of the crops and another questionnaire is sent by the end of the harvest period, in order to obtain the crop yields. The farmers are asked to send the questionnaires with information on all crops before 31 December. Intermediate information at different dates is collected by telephone.

Figure 12.1

France: Illustration of the TER-UTI Sample Selection Method
Aligned Systematic Sampling of 4 or 8 "Positions" in each Cell, and Selection of "Points"



Source: Amorich, S. et. al. (1997)

TER-UTI SURVEY
PHYSICAL AND FUNCTIONAL NOMENCLATURES

PHYSICAL CLASSIFICATION

Permanent waters and wetlands

11. Salt marshes and brackish swamps
12. Lakes, fresh water ponds
13. Rivers (including estuaries), canals
14. Swamps, humid zones (including marshes and peat bogs)
15. Glaciers and everlasting snow

Rock, pebbles, sand

16. Rocks, boulders
17. Dunes, beaches with sand or shingle

Wooded area

18. Deciduous trees
19. Coniferous trees
20. Deciduous forest evolving towards coniferous forest (changed in 1993)
21. Mixed forest
22. Copses
23. Scattered trees
24. Poplar groves
25. Poplar groves associated with other land use
26. Scattered poplars

Utilized agricultural area

27. Wheat
28. Barley
29. Oats
30. Maize
31. Rye
32. Maslin and other mixed cereals
33. Sugar beets
34. Textile crops
35. Sunflower
36. Rape
37. Other oilseeds
38. Other industrial crops
39. Potatoes
40. Peas and green peas
41. Broad beans and field beans
42. Other vegetables
43. Tree nurseries, flowers, ornamental plants
44. Fodder root crops
45. Other annual fodder crops
46. Leguminous grass or grazing
47. Mown pasture (changed in 1993)
48. Grazed pasture (changed in 1993)
49. Mountain grassland *(Alpages)*
50. Low-productivity grassland
51. Grassland under fruit-trees
52. Temporary fallow
53. Apricot trees
54. Cherry trees
55. Peach trees
56. Plum trees
57. Pear trees
58. Apple trees
59. Mixture of 53 through 58 ("six species mixed orchard")
60. Other fruit trees
61. Association of six species mixed orchar with non-fruit crops
62. Association of 60 with non-fruit crops
63. Vineyards
64. Association of vineyards and six species mixed orchard
65. Association of vineyards and other crops
66. Berry bushes
67. Vegetable kitchen gardens
68. Lawn and other grass area
69. Fallow land
70. Moors
71. Sclerophyllous vegetation *(garrigue, maquis,* merged with 70 in 1993)
72. Hedges

Artificialized areas

73. Non-surfaced roads
74. Building sites
75. Other land covers altered by extraction (quarries, etc.)
76. Other land covers altered by deposits (rubbish dump, etc.)
77. Graveyards
78. Urban waste grounds
79. Farmyards, annex farm buildings
80. Surfaced areas under trees
81. Surfaced areas without cover
82. Surfaced linear features (roads, etc.) under cover
83. Surfaced linear features (roads, etc.) without cover
84. Association of different non-agricultural elements
85. Low covered and closed buildings
86. Tall or medium-height buildings
87. Greenhouses
88. Covered non-closed buildings
89. Temporary buildings
90. Other industrial buildings
91. Disused buildings including ruins
99. Forbidden areas

FUNCTIONAL NOMENCLATURE

Primary production
1. Mining industry
2. Agriculture
3. Occasional agricultural production
4. Timber production
5. Fish production

Secondary production
6. Industry
7. Energy

Services and miscellaneous
8. Permanent road network
9. Railroad network
10. River or sea navigation
11. Air transportation
12. Other networks such as water distribution
13. Trade and services
14. Offices, administratrion and local communities
15. Military
16. Education and research
17. Culture
18. Places of worship
19. Social and health services
20. Sport and outdoor leisure activities
21. Individual dwellings
22. Collective dwellings
23. Rubbish dumps and other deposits
24. Protection of the environment
99. No use or temporary use

CHAPTER 13

ITALY

CROP PRODUCTION AGRICULTURAL SURVEY BASED ON AREA FRAME SAMPLING METHODS (1988-1997)

1. INTRODUCTION

This chapter describes the 1988-1997 Quarterly Italian Agricultural Survey based on Area Frame Sampling Methods promoted by the Ministry of Agriculture and carried out by the ITA Consortium (Consorzio Italiano per il Telerilevamento in Agricoltura). The survey objective is to obtain for each of the 20 Italian regions, and at the national level, crop area estimates for the main crops and crop production estimates for ten crops.

To date, the survey estimates have not been included as part of the Italian Statistical System (SISTAN) of the Italian Statistical Institute (ISTAT). The ISTAT does not use area sample methods for agricultural surveys.

2. SOME CHARACTERISTICS OF THE AGRICULTURAL SECTOR

Italy is administratively divided into 20 regions which are subdivided into 102 provinces and 8 101 municipalities.

Italian total area is 30 132 267 ha of which 26.96% (8 124 978 ha) is covered by arable land, 8.6% (2 603 219 ha) by permanent crops and 21.3% (6 418 116 ha) by forests. Cereals are the main crops, covering about 4 million ha. According to a 1993 ISTAT survey, there were 2 205 339 ha of wheat and 936 489 ha of maize; other widespread crops were: vineyards (900 574 ha), olive trees and orchards. There were also 7 591 327 heads of cattle and 8 396 340 hogs. According to the 1990 agricultural census, there were 3 023 344 holdings, corresponding to 22 702 355 ha of total area and about 15 million ha of utilized agricultural area (UAA); the average area of holdings is 7.5 ha with an average of 5.0 ha of UAA.

3. DEVELOPMENT OF AREA SAMPLE DESIGNS

Three main area sample designs have been used for the special purpose survey:

1988-1989: *Area Frame based on Square Segments*

After several years of experimentation, in 1988 the AGRIT survey programme to obtain area and production estimates for the main crops was established. The national area frame was based on area sampling methods with square segment methodology that was later adopted and developed by the Monitoring Agriculture with Remote Sensing (MARS) project of the Joint Research Centre (JRC) of the European Union.

1990-1991: ***<u>Area Frame based on Segments with Permanent Physical Boundaries. Strata and PSUs constructed using TM/Landsat Satellite Imagery</u>***

The area sample design with square segments was replaced by an area sample design of segments with recognizable physical boundaries - the main area sample method described in Volume I - in view of the relative advantages: to improve efficiency and reduce non-sampling errors due to incorrect identification of segments on the ground, which are frequent in the case of segments with non-recognizable physical boundaries.

In 1990, the area frame based on segments with permanent recognizable physical boundaries was constructed. Segment size was 100 ha for hilly areas and 50 ha for plains. The cartographic materials used for area frame construction were: TM/Landsat satellite images and cartography, scale 1:100 000 for delineating counting units (CUs), topographic charts in scale 1:50 000 and aerial photographs for the other phases of the work.

1992-1997: ***<u>Area Frame based on Segments with Permanent Physical Boundaries. Strata and PSUs constructed using SPOT Satellite Imagery</u>***

From 1992, a new area sampling frame has been used, with the land use stratification carried out at scale 1:25 000. A set of SPOT images covering the Italian territory become available; consequently the area frame was changed again. At present, estimates are provided at the regional level and their coefficients of variation are less than 10% for the main crop areas in each region.

This area frame, of segments with recognizable physical boundaries, promoted the experimentation of a general purpose agricultural survey in a few Italian regions, as described in section 5.

The 1996 area sample design will be described in the following sections.

4. SURVEY DESIGN AND AREA FRAME CONSTRUCTION

4.1 <u>The Area Sample Design</u>

The area sample design is analogous to the main one described in detail in Vólume I. The sample segments are land areas with identifiable physical boundaries. The total sample in 1996 included 5 500 segments. The sampling rate is currently about 1.5% of the Italian agricultural area.

The survey has been conducted four times a year to collect information on winter and summer crops. From 1996, areas of winter cereals, soybean and sugar beet, at the national level, will also be estimated in April on the basis of a survey conducted on a 20% Simple Random Sample (SRS) subsample of segments in each region.

4.2 Survey Variables

The survey variables are the *area* of main crops and the *production* of the following ten crops: soft wheat, durum wheat, barley, grain maize, sunflower, soybean, rapeseed, sugar beet, tobacco and tomato.

4.3 Land Use Stratification. Survey Area

The following land use strata were identified through visual interpretation, on multispectral SPOT images enlarged to scale 1:25 000, acquired in the summers of 1988 and 1989. In fact, summer time is the best period to analyse agricultural features from the images.

Land use strata definitions

Sampled Strata (Survey Area)

2.1 More than 75% of cultivated area; little or no presence of permanent crops.
2.2 More than 75% of cultivated area; more than 30% under permanent crops (orchards, vineyards olive trees).
3.1 Between 50% and 75% of cultivated area; little or no presence of permanent crops.
3.2 Between 50% and 75% of cultivated area; more than 30% under permanent crops (orchards, vineyards, olive trees).
4 Less than 50% of cultivated area.

Non-Sampled Strata

5 Permanent grassland and pasture.
6 Forest land and woods.
7 Urban areas (industrial, residential, commercial).
8 Water bodies and permanent ice.
9 Areas without vegetation.

The survey area consists of the strata numbered 2.1, 2.2, 3.1, 3.2 and 4, which cover a total of 135 130 km^2.

Stratum 5 was delineated during the photo interpretation, but is not included in the survey area. Strata 6, 7, 8 and 9, corresponding to about one-third of the national territory, are excluded from the survey area because they are considered non-agricultural. The non-sampled strata cover an area of 165 810 km^2.

4.4 Counting Units and Segments

Segments. The segment target size for all strata was 50 ha, with a tolerance of ±5%.

Counting Units. Each land use stratum was subdivided into counting units (CUs). The target size of the CUs in a given stratum was a multiple of 50 ha (the segment target size), with a tolerance of plus or minus 5%; while their maximum size was determined to be 650 ha.

4.5 Sample Size and Allocation. Area Frame Parameters

As already mentioned, in most regions the sampling rate is about 1.5% of the survey area, as defined in 4.3 above. The total sample size of 5 500 segments was determined with the aim of obtaining coefficients of variation of less than 10% for the main crop areas in each Italian region.

Neyman allocation was applied separately in each region for each variable; a compromise for the different optimum segment target size was adopted in each region on the basis of Chatterjie or Yates methods. Tables 13.1 and 13.2 show the segment allocation in the Emilia Romagna and Marche regions in 1992. The table below shows the total and sampled areas, and the number of sampled segments in the 1995 survey for three regions.

Sample size in three regions

Regions	Total area (ha)	Sampled area in 1995 (ha)	Selected segments in 1993 and 1995
Marche	973 187	761 976	410
Campania	1 354 003	805 267	620
Veneto	1 821 156	1 083 819	750

4.6 Data Collection

4.6.1 Data Collection Materials

For each segment, the enumerator uses an enlargement of an aerial photograph in the scale of 1:4 000 to 1:7 000,or a topographic map or an orthophoto in scale 1:10 000. The aerial photo enlargements used were obtained from 1992 and 1993 aerial photographs with a scale of 1:40 000.

Updated orthophotos are the best cartographic material to use for data collection.

The use of topographic maps, *vis-à-vis* aerial photography, results in lower costs for segment allocation and enlargement reproduction(a laser photocopier suffices), makes digitalization of segments easier and requires fewer ground control points. Topographic maps are also easier to use in the plains, where many cartographic features recognizable on the ground are available. However, topographic maps are generally old and outdated. Aerial photographs require higher costs of enlargement reproduction (photographic enlargements have to be produced from scale 1:40 000 to scale 1:5 000), and more work for the geometric correction of segments but, on the other hand, they are generally more up to date.

4.6.2 Data Collection Procedures

The survey is carried out four times a year with a total of 120 enumerators who visit the sample segments but do not perform direct farmer interviews. During the field data collection, the enumerators delineate on the photo enlargement the boundaries of the segments and fields and other areas in the segments.

Subjective crop yield estimation is carried out twice a year by some of the enumerators (who are also agronomists) in two fields in each sampled segment for each of the ten crops. Maize yield subjective estimates are made a third time in some regions.

During the *first annual survey round*, all land uses in the segments are mapped with the aim of collecting data to estimate the areas covered by winter crops. All areas larger than 600 m^2 and all linear features longer than 20 m are delineated by the enumerator. During the *second annual survey round*, the land use of the segments is updated to estimate the area of winter crops taking into consideration crops seeded in the spring. Yields are also estimated for appropriate crops. The *third annual survey round* is conducted in order to estimate the areas under summer crops. The *fourth annual survey round* is conducted in segments with maize in order to distinguish between maize grown for grain or for silage. An update of the yield subjective estimate is also undertaken.

A field quality control operation is carried out by a commission instituted by the Ministry on 2% of the sample segments. The field boundaries and the land use are checked. Independent quality checks are performed by the ITA Consortium.

4.7 Data Processing. Digitization

The maps with the land use for surveyed segments are digitized with digitizing tables or a telecamera and geometrically corrected through an automatic procedure in order to calculate field areas and to allow the overlapping of remote sensing images.

4.8 Estimation Methods and their Precision

4.8.1 Crop Area Estimation

Crop area estimates are obtained for each region; estimates are then aggregated at the national level.

Whenever possible, crop area estimates are obtained by using an estimator that combines field survey data with data derived from the analysis of TM/Landsat satellite data.

The crop area estimates obtained by classifying satellite data are used as an auxiliary variable to adjust, with a regression estimator, the area estimate of each crop based on field survey data collected on sampled segments. In fact, the area estimate of each crop based on surveyed segments is modified on the basis of the difference between the average number of pixels classified to the crop in the population and the average number of pixels classified for the crop in the sample.

Direct estimation of some crop areas using TM/Landsat Satellite Images

The Italian territory is covered by 22 TM/Landsat images. A complete coverage is acquired approximately in the periods 15 April to 30 May and 5 July to 31 August, which are the proper acquisition dates in view of the crop growing periods. Sometimes large parts of the images are covered by clouds, in which circumstances images of dates near the ones described are acquired. When the acquisition dates are distant from the optimum acquisition periods, it is difficult or often impossible to discriminate between the different crops.

Images are geometrically corrected through 25 ground control points, obtaining an average error of about one pixel. A combination of TM bands 3 (red), 4 (near infrared) and 5 (middle infrared) is used. Then spectral signatures of the different crops are determined through an automatic procedure whose inputs are the number of clusters and the minimum number of pixels per cluster. Finally, all the pixels in each stratum of each image are classified through a maximum likelihood algorithm; thus, the number of pixels classified for each crop is calculated for each segment in the population in each stratum.

The regression estimator

The average area per segment of each crop is adjusted with the following regression estimator:

$$\bar{y}_{reg} = \bar{y} + b(\bar{X} - \bar{x});$$

where:

$\bar{y}$ = average crop area per segment calculated on the basis of field survey data;

$\bar{y}_{reg}$ = average crop area per segment adjusted by means of regression;

b = angular coefficient of the regression line;

$\bar{X}$ = average number of crop pixels per segment in the population, determined by classification of satellite data;

$\bar{x}$ = average number of crop pixels per sampled segment.

The variance of $\bar{y}_{reg}$ is smaller than the variance of $\bar{y}$ by $(1 - r^2)$, where r is the correlation between area effectively covered by the crop and the number of pixels classified for the crop in each sampled segment. Thus, the higher the correlation between field and satellite data, the higher is the precision of the regression estimator. On the other side, the regression estimator is biased and should not be used when a small number of segments is present in a stratum or the classification of remote sensing data is not reliable; thus the difference of b and r from zero is tested.

A study is being undertaken to assess the convenience (cost-effectiveness) of using the estimation procedure described that requires the acquisition of semestral satellite images and the utilization of the regression estimator.

4.8.2 Production Estimation

As already mentioned, survey yield estimates are obtained for ten crops, namely: soft wheat, durum wheat, barley, grain maize, sunflower, soybean, rapeseed, sugar beet, tomato and tobacco. These estimates are combined with crop yields given by an agro-meteorological model named SAM used also to forecast crop yields. These crop yield estimates are then used with the survey crop area estimates to obtain the production estimates.

SAM is partly a growth model and partly a statistical model; in fact, it includes, in a regressive type structure, some variables which have a precise significance from the agronomic viewpoint. These characteristics allow the use of SAM for large areas and, at the same time, ensure interpretation of the agronomic reality. Considered variables intend to represent the water and thermal stresses the plants may suffer during their growth phases. The influence of stresses, both thermal and water type, on the productive capacity varies during the growth period; consequently, different coefficients are estimated for the two kinds of stress in the different phenological phases.

Every ten days, from March to October, a statistical bulletin is delivered to the Ministry of Agriculture, and the same data and agrometeorological information are supplied by remote connection to government organizations.

4.9 Some Survey Results and their Precision

The coefficients of variation of the main crop area estimates are in most cases less than 10 %. Coefficients of variation are not calculated for yield or production estimates; in fact, production estimation is calculated by the area estimates multiplied by yield estimates. Yield estimates are obtained by using the agrometeorological model SAM, whose precision is very difficult to evaluate.

5. GENERAL PURPOSE AGRICULTURAL SURVEYS BASED ON AREA OR MULTIPLE FRAME SAMPLING METHODS (1992-1997)

5.1 Survey Area and Survey Variables

This section presents the general purpose agricultural surveys based on area sampling methods, conducted in different regions of Italy. These surveys, using the area sample frame described in previous sections but with higher sampling rates, were also carried out by the ITA Consortium. Their purpose was to obtain estimates of UAA, area and production of ten main crops, type of management, area where cultivation has been discontinued, irrigated area, livestock numbers, labour, nitrogen use by crop and structural characteristics of the farms. Crop area estimates are obtained at the province level.

The surveys were conducted in the Marche region in 1992, in Campania, Marche and Veneto in 1993, in all Italian territories in 1994, in Campania, Marche and Veneto in 1995 and in the Basilicata and Lombardia regions in 1996.

5.2 Pilot Tests

Pilot surveys were carried out using area sample segments of two provinces (Mantova in 1990 and Foggia in 1991), one in the north and the other in the south of Italy. The farms were those partially or totally included in the surveyed segments. Open, closed and weighted segment estimators were compared and a random sampling of various subsets of farms was simulated. The weighted segment estimator proved to be the most stable and efficient in the study areas. Another result was that information collected in different farms in the same segment tend to be redundant for many variables; thus a subset of farmers can be interviewed with little loss of information.

5.3 Overall Sampling Design

The two-stage area sample design of the general purpose surveys is based on the area sample design described in previous sections. The general purpose survey design involves substratification and a higher sampling rate than the crop and production area sample design. In fact, estimates are obtained for each province, a smaller administrative level than the region.

5.4 Substratification based on Clusters of 1990 Census Data by Municipality. Construction of PSUs

The substratification was based on a cluster of 1990 census data at the *municipality* level. The percentages of UAA under the main crops were calculated for each municipality. Municipalities in a given survey *region* (for a given year) were clustered on the basis of the minimization of within variance and maximization of variance between clusters.

A check was done on the spatial distribution of municipalities in the same cluster and the procedure was repeated with little modification if municipalities in the same cluster were scattered over the region.

The map of clusters was then overlapped on the map including the land use strata. Finally, PSUs were assigned to different strata on the basis of prevalence; thus, all PSUs are included in only one stratum.

5.5 Sample Size

Sample size was determined as a function of the variance of the area of main crops in each province; in particular, the sampling rate in each province was determined with the aim of achieving a coefficient of variation less than or equal to 5% for the main crop area estimates in each province. For some other crops, covering an area of 10 000 ha or more, a coefficient of variation less than or equal to 10% was established. Table 13.2 shows the segment allocation in the Marche region in 1992.

5.6 Sample Selection

In each sample segment, farms are selected by point sampling. A regular grid is overlaid on each segment and (depending on the characteristics of the region) five or six couples of coordinates are selected. The coordinates identify points on the ground. If a point falls on the utilized agricultural area (UAA) of a farm, the farm is selected; otherwise, if a point falls on a non-UAA, a zero value for all variables is attributed to a fictitious farm corresponding to that point. Farms are selected with replacement and with probability proportional to the fraction of their UAA included in each sample segment.

For the 1993 survey, five points were selected in the sample segments of the regions of Marche and Veneto and six points in each sample segment in the Campania region. Thus, the number of different selected farms was ≤ 5 in Marche and Veneto and ≤ 6 in Campania, since some points could fall on the same farm.

The following table, for the 1993 survey covering three regions, shows the sample segments, the total number of selected points, the number of points falling on utilized agricultural area (UAA) and the points corresponding to respondent farmers.

1993 Farm Survey

Regions	Selected segments	Selected points	Points falling on UAA	Points corresponding to respondent farmers (selected farms)
Campania	620	3 720	2 821	1 205
Marche	410	2 050	1 423	1 352
Veneto	750	3 750	2 471	1 743
TOTAL	1 780	9 520	6 715	4 300

5.7 Estimation Methods

For the estimation of a variable total, a weighted segment approach, in two-stage sampling, is used. In each sample farm of a segment, the value of the variable attributed to the tract is a function of the percentage of UAA of the total farm in the tract. The estimation method for the two-stage design is the following:

If W_{ik} is an additive variable on the whole farm k of the segment i, then an estimator of the total of variable X in stratum W_h is:

$$\hat{X}_h = \frac{N_h}{n_h} \sum_{i=1}^{n_h} \frac{1}{K_i} \sum_{k=1}^{K_i} \frac{T_{ik}}{A_{ik}} W_{ik} \frac{D_i}{T_{ik}} = \frac{N_h}{n_h} \sum_{i=1}^{n_h} \frac{D_i}{K_i} \sum_{k=1}^{K_i} \frac{W_{ik}}{A_{ik}}$$

where:

N_h is the total number of segments in stratum W_h (the population W of segments is divided into strata W_h, h=1, ..., H).

n_h is the number of sample segments in stratum W_h.
K_i is the number of sampled points in the i^{th} segment,
A_{ik} is the UAA of the entire k^{th} farm, inside and outside the i^{th}segment.
T_{ik} is the UAA of the tract corresponding to the k^{th} farm of the i^{th} segment.
D_i is the area of i^{th}segment

This means that, even if the second stage sampling unit is the UAA of the farm included in the segment, it is not necessary to know its area nor X_{ik}, but just the global information about the farm (W_{ik}). Its variance in stratum h can be estimated as (cf. Cochran, 1977, section 11.6):

$$\hat{V}(\hat{X}_h)=\frac{N_h^2}{n_h}\left(1-\frac{n_h}{N_h}\right)\sum_{i=1}^{n_h}\frac{(\hat{X}_i-X_h)^2}{n_h-1}+\frac{N_h}{n_h}\sum_{i=1}^{n_h}\frac{1}{K_i(K_i-1)}\sum_{k=1}^{K_i}\left(\frac{W_{ik}D_i}{A_{ik}}-\hat{X}_i\right)^2$$

The estimates for the total are:

$$\hat{X}=\sum_{h=1}^{H} X_h \quad \text{and} \quad \hat{V}(\hat{X})=\sum_{h=1}^{H}\hat{V}(\hat{X}_h)$$

For the 1995 and 1996 surveys, multiple frame designs were adopted. In Campania and Veneto regions, the list sample component includes the largest farms and was used to complement the area sample component. In the Marche region, the list sample includes all farms.

5.8 Data Collection Procedure

The enumerators conduct personal interviews with farmers once a year, generally in December. Data concerning livestock refer to the survey period, while, for all other variables, the reference period was the previous agricultural year, from 1 December to 30 November.

5.9 Missing Data

Missing data are a frequent problem which arises for several reasons. For instance, sometimes it is very difficult to find the farmer whose land corresponds to a selected point on the ground, or the farmer refuses to cooperate and the enumerator cannot complete the questionnaire.

The imputation procedure adopted for missing values was the following: if data for a farm are missing, a value for each variable is assigned on the basis of the weighted average of the farms in the segment. If there are no data for any farm of a given segment, then for each variable the average of the stratum is assigned to the segment considering its UAA.

For the estimation of the variance, the sample size is given by the number of respondents.

If missing data are ignored, the estimates would have a significantly negative bias since missing data could only belong to UAA.

Table 13.1
Italy: Emilia Romagna region area frame parameters

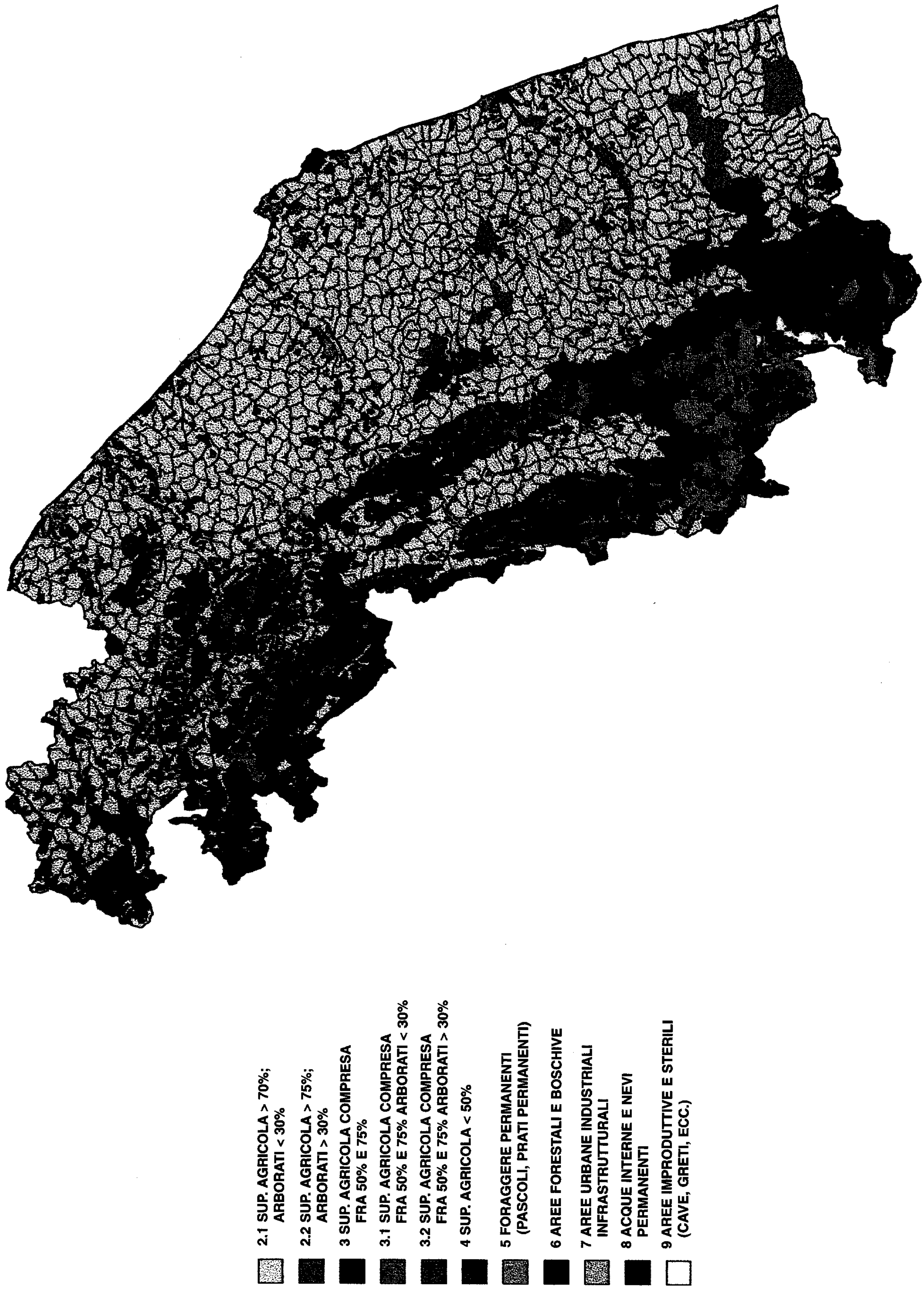
Emilia Romagna region - 1992						
Stratum	Total area (km^2)	Target segment size (km^2)	Total number of segments	Number of sample segments	Average expansion factor	Stratum definition
2.1	11 729	0.5	23 458	295	79.52	More than 75% of cultivated area; no or little presence of permanent crops
2.2	2 870	0.5	574	80	71.75	More than 75% of cultivated area; permanent crops more than 30%
3.1 +3.2+4	2 617	0.5	5 234	45	116.31	From 50% to 75% of cultivated area Less than 50% of cultivated area
5	97	0.5	194	0	-	Permanent grassland and pasture
6	3 645	0.5	7 290	0	-	Forest land and woods
7	570	0.5	1 140	0	-	Urban areas
8	240	0.5	480	0	-	Water and permanent ice
9	320	0.5	640	0	-	Areas without vegetation

Table 13.2
Italy: Marche region area frame parameters

Stratum	Substrata	Total area (km^2)	Target segment size (km^2)	Total number of segments	Number of sample segments	Average expansion factor	Stratum definition
2.1+2.2	A	2 246	0.5	4 492	100	44.92	More than 75% cultivated; prevalence of durum wheat and sunflower
2.1+2.2	B	1 298	0.5	2 596	40	64.90	More than 75% cultivated; prevalence of durum wheat, soft wheat and sugar beet
2.1+2.2	C	831	0.5	1 662	45	36.93	More than 75% cultivated; prevalence of soft wheat and fodder from arable land
2.1+2.2	D	553	0.5	1 106	25	44.24	More than 75% cultivated; prevalence of permanent crops: vineyards and olive trees
2.1+2.2	E	1 248	0.5	2 496	30	83.20	More than 75% cultivated; prevalence of fallow and permanent grassland
3.1+3.2+4	A+B+C+D	518	0.5	1 036	25	41.44	From 50 to 75% of cultivated area; prevalence of arable land
3.1+3.2+4	E	927	0.5	1 854	45	41.20	From 50 to 75% of cultivated area; prevalence of fallow and permanent grassland
5		474	0.5	948	0	-	Permanent grassland and pasture
6		1 388	0.5	2 776	0	-	Forest land and woods
7		163	0.5	326	0	-	Urban areas
8		32	0.5	64	0	-	Water and permanent ice
9		56	0.5	112	0	-	Areas without vegetation

Figure 13.1

Italian area sampling frame: Marche region stratification and PSUs



CHAPTER 14

SPAIN

CROP AREA AGRICULTURAL SURVEY PROGRAMME BASED ON AREA FRAME SAMPLING METHODS (1988-1997)

1. INTRODUCTION

Before 1988, almost all agricultural statistics in Spain were based on censuses, list sample surveys, non-probability surveys, expert opinions or on administrative data. Livestock surveys are list sample probability surveys. The remaining official agricultural statistics are still based on the non-probability methods mentioned above.

The European Commission requires that member states make objective estimates of area and yield of main crops with a certain precision. For Spain, the accepted coefficients of variation (CVs) are the following: 1% for area of cereals, 2% for production of cereals, 1% for vineyards, while, for other main crops, a 3% CV is required.

To meet these requirements, in 1988 the Ministry of Agriculture Statistics Office began the development of area sample surveys for the estimation of the main crop areas in the Castilla y León region, and was directly linked with the Monitoring Agriculture with Remote Sensing (MARS) project of the Joint Research Centre (JRC) of the European Union, at Ispra, Italy. The development of the area sample survey has also been linked to tests of the use of high-resolution satellite images to improve the accuracy of the estimates by a regression correction on classified satellite images. This regression correction procedure was too expensive for practical application.

Since 1997, the survey area covers about 470 000 km^2, including 13 of the 17 *administrative regions* (autonomous communities). The area sample survey provide annual estimates of areas with wheat, barley, oats, rye, maize, potatoes, sugar beet, sunflower, alfalfa, olive trees, vineyards and fallow.

The area sampling frame is not stratified. Multiple frame sampling (combination of an area sample and list sample) has not been used in Spanish surveys to date.

2. SURVEY AREA

Spain is divided into the following 17 administrative regions: Andalucía, Aragón, Asturias, Baleares, Canarias, Cantabria, Castilla-La Mancha, Castilla y León, Cataluña, Extremadura, Galicia, La Rioja, Madrid, Murcia, Navarra, País Vasco and Valencia.

As already mentioned, the development of area sample surveys for the estimation of the main crop areas started in 1988, in the Castilla y León region. The area sample survey was then extended to the Madrid and Castilla-La Mancha regions; finally, as of 1997, the survey area now covers about 470 000 km^2, including the following 13 regions: Andalucía, Aragón, Baleares, Castilla y León, Castilla-La Mancha, Cataluña, Extremadura, Galicia, La Rioja, Madrid, Murcia, Navarra and Valencia. In Navarra and Valencia regions, the survey is conducted by the regional administrations.

Main Crop Areas and Land Use in Spain. According to the 1989 National Agricultural Census, there were 24 740 506 ha of utilized agricultural area (UAA) in the farms. Around 38.4% of the area is covered by forest and scrub; 57.6% of the total surface is UAA and 4% is covered by other land uses. Cereals cover 29.8% of the UAA, while other annual crops cover 19.1%. Other main agricultural land uses are: orchards and citrus correspond to 4.9%, olive trees 7.2% and vineyards 4.4%. Meadows and rough pasture correspond to the remaining 34.3%.

3. GENERAL SURVEY DESIGN

3.1 The Area Frame of Square Segments

The area frame carried out in the above regions of Spain makes use of the Universal Transverse Mercator (UTM) projection of 1 x 1 km squares that covers the survey area and is printed on the standard 1:50 000 scale topographic maps. A pilot survey in Navarra compared the results of an area sample whose segments (sample units) had physical boundaries with an area sample where the segments were squares. In this case, the precision was found to be about equal but the cost of constructing a frame with physical boundaries was significantly higher. With this in mind, it was decided to base the frame on the UTM grid. Further investigation showed that using the 1 x 1 km squares of the grid for segments would result in an excessive number of parcels (tracts) in each segment. Experience in a number of countries working with the MARS project tended to indicate that non-sampling errors can be maintained within tolerable limits with a segment containing 15 to 25 tracts. In the survey area, this number of tracts would correspond to a segment having an area between 35 and 58 ha.

A *segment* (sampling unit) is a square of 700 x 700 m with an area of 49 ha inserted into the 1 x 1 km UTM grid units and is expected to contain about 20 tracts. For sampling purposes, the 1 x 1 km units were grouped in blocks of 10 x 10 km (100 km^2). Three units were selected in each block.

Some problems arise with squares that straddle the border of the target area. An approximation of the region is made by dropping the small pieces and keeping the large pieces giving them the same weight as if they were full squares.

Stratification of the sampling frame is not used.

3.2 The Sample Design. Sample Selection

For the purpose of sample selection, the survey area was divided into 10 x10 km *blocks* using the 1 x 1 km UTM grid squares over the survey area. In each of these blocks three segments are selected following a pattern that is repeated across all the blocks. A distance threshold of 3 km was chosen so that the distance between pairs of sample segments should be no less than 3 km.

A systematic aligned sampling selection procedure by square blocks with three replicates and a distance threshold was used.

The sampling design consists of a cluster sampling procedure where a cluster is the set of all possible segments with the same relative position in each block. The sample consists of three clusters drawn systematically out of a universe of 100 clusters with an incomplete observation: 49 ha (segment size) are observed out of 100 ha (UTM cell).

The segments are positioned in the southwestern corner of the 1 x 1 km UTM grid squares. The remaining 51 ha in each grid square do not form a part of the sampled population and, therefore, cannot be selected in the sample. This is not expected to introduce a bias in the estimation, since the location of a point at a corner of a UTM cell is assumed to be independent of the land use and both populations have similar spatial distribution.

In 1997, 14 246 segments were selected in the survey area. The sampling rate is around 1.5%.

The sampling procedure is as follows:

- Select, with equal probability and without replacement, 3 of the 100 1 x 1 km squares that make up the 10 x 10 km block. This is done by choosing a random number between 1 and 10 along the x and y axes of each block and locating the selected square by counting out from each axis the spaces indicated by the random numbers. The segments occupy positions 2-6 (abscissa 2 and ordinate 6 in relation to the southwestern corner), 3-2 and 6-7.
- Include in the sample the three squares from *each block in the population* that occupy the same position in the block as those selected above (cf. Figure 14.1).

This procedure for selecting the sample can be considered systematic with three random starting points.

3.3 Area Sample Estimators and their Precision

The application of classical formulae for cluster sampling gives an unbiased estimate of the total area but the variance estimator is slightly biased because it considers neither the distance threshold (positive bias) nor the fact that only a part (49 ha) of each selected UTM cell (100 ha) is observed, as already explained (negative bias). The estimator of the variance is, however, very unstable because of the very small sample size (n = 3). For this reason, a more biased but more stable estimate of the variance is used by considering the sample as a random sample of segments instead of as a sample of clusters. Assuming that the spatial autocorrelation is positive, there is probably a positive bias of the variance, i.e. the calculated variance is larger than the error actually made.

An unbiased estimation of the proportion of crop c in a region is given by:

$$\hat{\bar{y}} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

where:

y_i is the proportion of the segment *i* covered by crop *c*;
N is the total number of segments in the population;
n is the number of segments in the sample.

The total area of crop *c* is given by:

$$\hat{z} - D\hat{\bar{y}}$$

where D is the total area of the region.

$$V(\hat{\bar{y}}) = (1 - \frac{n}{N}) \frac{1}{n(n-1)} \sum_{i=1}^{n} (y_i - \bar{y})^2 \quad \text{and} \quad V(\hat{z}) = D^2\, V(\hat{\bar{y}})$$

3.4 Data Collection

The enumerator outlines all fields within the segment on a transparency placed over a segment photo. Field and segment outlines are then transferred to a larger transparent sheet, scanned and digitized. Enumerators also record, using a grid point, the count of points for each field so that a quick estimate of area can easily be made. This also allows a check of the digitizing operation.

During field data collection, *no interviews with holders or farmers are undertaken*. In each sample segment the area of each main crop is measured by the enumerators during the annual field data collection. However, in order to reduce costs, ground visits and measurements for sample segments without arable land are replaced by a photo-interpretation of aerial photographs.

Besides the 1:50 000 scale topographic maps, enumerators use aerial photographs (contact prints) at a scale of around 1:20 000, where sampled segments are drawn as well. The maps and contact prints are used to locate the segment on the ground.

When a segment has been located, an enlarged aerial photo in a scale of 1:5 000 showing the segment boundaries is used to guide the data collection (cf. Figure 14.2). A transparency is overlaid on the enlargement and centred on the segment. The segment boundaries are drawn on the overlay. During the fieldwork, the boundaries of the fields are drawn using a black waterproof felt-tip pen, and field numbers are written with a red pen on the overlay (cf. Figure 14.3). The colour difference allows for line filtering when digitizing. Land use and field numbers are recorded on a form together with a rough area measurement made by counting points on a point grid that fits over the segment (a 50 m step grid). Linear elements of the landscape with a width less than 15 m are represented as a line with no area.

Whether or not the field is irrigated is recorded. Nothing is recorded about the holder or operator of the fields.

About 40 enumerators, generally experts collaborating with agricultural insurance companies, are involved in the field data collection. Each enumerator visits about 30 segments and needs about three hours for the ground work of each segment. About 2% of the segments are surveyed independently by a supervisor to assess the quality of the ground work. The fieldwork is not carried out for all the 14 246 sample segments. In fact, about 6 000 segments (42%) which do not include any arable land are not surveyed every year, but simply photo-interpreted. However, once every four or five years these segments are visited, unless they are inaccessible.

Timing of the surveys must be done so that all crops in a given segment can be identified and crops such as barley and wheat can be distinguished from each other. Separating barley and wheat has presented a problem in previous surveys.

3.5 Data Processing

Segments are drawn more cleanly on Din-A0 sheets (35 segments per sheet) and digitized with a scanner. Data are transformed into vector mode using Arc-Info PC. Digitized area and land use code for each field are stored in DBase together with the enumerator's comments and the number of points of a grid counted by the enumerator (this provides a way to detect errors).

3.6 Survey Results

The coefficient of variation of area estimates varies with the size of the sample and the relative importance of the crop. For example, for a large region (80 000 km^2 with a sample of 2 400 segments), the coefficient of variation is about 2% for main crops that cover more than 10% of the region, and about 6% for crops that cover 1% of the region.

Results are generally coherent with subjective estimates, except when fast changes take place: local experts tend to be prudent and give a reduced estimate of the change (cf. Table 14.1).

Table 14.1

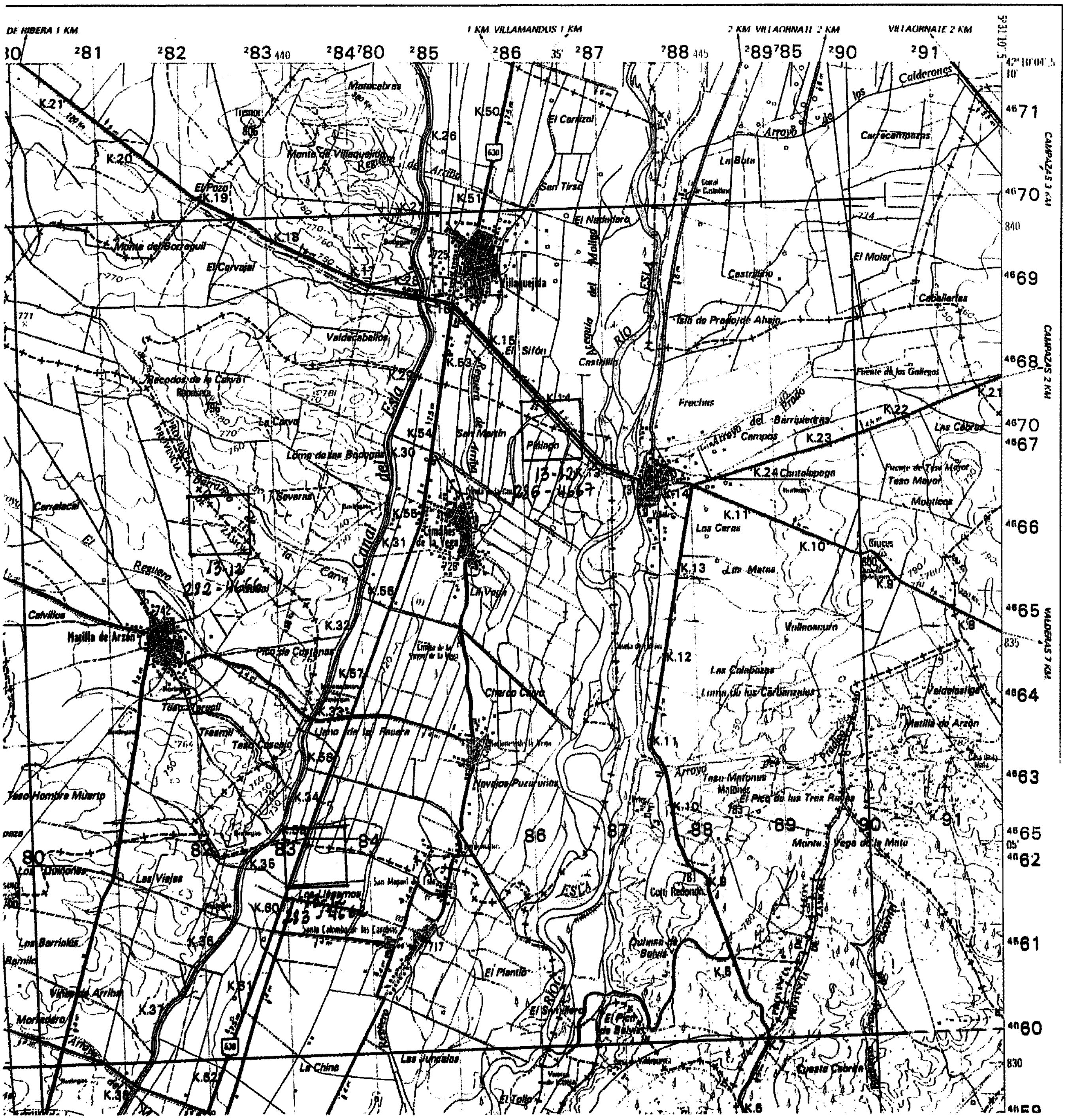
Comparison between segments survey estimates and traditional procedure (adding subjective estimates of local experts) in three regions in 1991

Regions and crops	Area sample survey (ha)	Subjective estimates (ha)
CASTILLA Y LEON		
Wheat	602 657	702 877
Barley	1 610 577	1.563 224
Oats	57 740	61 444
Rye	102 324	115 949
Maize	57 844	47 053
Potatoes	34 062	37 534
Sugar beet	81 493	83 150
Sunflower	186 503	159 466
Alfalfa	90 490	83 662
Fallow	758 241	886 356
Vineyards	79 435	71 217
Olive trees	6 292	10 711
MADRID		
Wheat	21 215	35 050
Barley	77 841	50 000
Oats	3 041	1 500
Rye	512	900
Maize	8 926	8 500
Potatoes	448	3 600
Sugar beet	273	70
Sunflower	2 166	2 800
Alfalfa	1 503	1 980
Fallow	44 674	93 016
Vineyards	24 211	25 063
Olive trees	19 565	22 163
CASTILLA-LA		
MANCHA	281 646	352 100
Wheat	1 222 153	1 167 700
Barley	111 259	121 300
Oats	13 259	23 700
Rye	81 300	79 812
Maize	6 873	11 956
Potatoes	2 590	8 213
Sugar beet	220 075	252 125
Sunflower	21 898	33 450
Alfalfa	753 804	1 021 398
Fallow	696 252	708 141
Vineyards	284 302	1 276 127
Olive trees		

Note: The data concerning fallow, vineyards and olive trees refer to 1990.
Source: MAPA.

Figure 14.1

Fixed position of the three selected segments in each block



Source: Ambrosio and Gallego (1994)

Figure 14.2

Aerial photo enlargement (scale 1:5 000) of a segment (code 13-12-283-4662) used for field data collection

Figure 14.3

Overlay of the aerial photo enlargement (scale 1:5 000) of a segment (code 13-12-283-4662) used for field data collection, with field boundaries

Source: Ambrosio and Gallego (1994).

INDEX OF TECHNICAL TERMS

BIBLIOGRAPHY

I. GENERAL SURVEY SAMPLING METHODS

Cochran, W.G. 1977. *Sampling techniques*. Third edition. John Wiley and Sons.

Deming, W. Edwards. 1960. *Sample design in business research*. John Wiley and Sons.

Hansen, M.H., Hurwitz, W.N. & Madow, W.G. 1953. *Sample survey methods and theory*. Vols I-II. John Wiley and Sons.

Hansen, M.H., Hurwitz, W.N. & Jabine, T.B. 1963. *The use of imperfect lists for probability sampling at the US Bureau of the Census*. Invited Paper, 34th session ISI, Ottawa.

Kish, L. 1965. *Survey sampling*. John Wiley and Sons.

Murthy, M.N. 1967. *Sampling theory and methods*. Calcutta Statistical Publishing Society.

Raj, D. 1972. *The design of sample surveys*. McGraw-Hill.

Sukhatme, P.V & Sukhatme, B.V. 1970. *Sampling theory of surveys with applications*. Second edition. London, Asia Publication House.

Wolter, K.M. 1985. *Introduction to variance estimation*. Springer-Verlag.

Yates, F. 1981. *Sampling methods for censuses and surveys*. Fourth edition. London, Griffin & Co.

II. AGRICULTURAL SURVEY SAMPLING METHODS

Ambrosio Flores, L. 1987. *"Estimación de rendimientos de cosechas de cereales de invierno por técnicas de muestreo. Aplicación a la tasación de daños"*. Investigación Agraria. *Economía*, 2 (1): 49-69.

Ambrosio Flores, L., Alonso, R. & Villa, A. 1993. *"Estimación de superficies cultivadas por muestreo de áreas y teledetección. Precisión relativa". Estadística Española*, 35 (132): 91-103.

Ambrosio Flores, L. & Gallego Pinilla, J. 1994. *"Encuestas agrícolas en España basadas en un marco de áreas"*. Case study prepared for FAO.

Amorich, S., Poiret, M., Wullems, M.-A., Palacio-Rabaud, V. & Giralt, A. 1997. *"Précision de l'enquête: utilisation du territoire"*. Service central des enquêtes et des études statistiques (SCEES), Ministry of Agriculture and Fisheries, Paris.

Beliu, Z. & Gjika, A. 1997. *Sampling frame surveys in Albania*. Service of Statistics and Information, Tirana, Albania. English translation by A. Karaj.

Beller, N.D. 1995. *Applied sampling techniques. Text and study sssignments.* Ministry of Agriculture and Food. Government of Albania. Description of Albanian Surveys, p. 185-196.

Bouzaffour, S. 1994. *"Maroc: étude de cas national sur les bases de sondage aréolaires/de listes pour les enquêtes agricoles".* National case study prepared for FAO.

Bush, J. and House, C. 1993. *The area frame: a sampling base for establishment surveys.* Proceedings of the International Conference of Establishment Surveys, Buffalo, NY, USA, 27-30 June, p. 335-344.

Caudill, C. 1960s. *Joint use of different sampling frames.* Statistical Reporting Service Staff Report, Washington, DC, USDA.

Candill, C. 1977. *Estimation of agricultural production by sample surveys, the US experience.* International Statistical Institute, 41st Session, New Delhi, India.

Clark, M.S. & Geuder, J. 1986. *Use of digital LANDSAT data to stratify area sampling frames for agricultural surveys.* Statistical Reporting Service Staff Report, Washington, DC, USDA.

Cotter, J.J. & Nealon, J. 1987. *Area frame design for agricultural surveys.* National Agricultural Statistics Service, Washington, DC, USDA.

Cotter, J.J. & Tomczak, C. M. 1994. *An image analysis system to develop area sampling frames for agricultural surveys.* Photogrammetric Eng. and Remote Sensing, 60 (3): 299-306 (March 1994).

Davidson, G. 1984. *1983 National Farm Survey.* Statistics Canada Methodology Branch Working Paper No. IASM-85-084E.

Dodson, Y.H. 1994. *The stratification process.* National Agricultural Statistics Service, Washington, DC, USDA.

Dodson, Y.H. 1966. *Landsat analysis.* National Agricultural Statistics Service, Washington, DC, USDA.

Dunkerley, C. & Huddleston, H.F. 1974. *A methodological report on: agricultural statistics in the Dominican Republic.* Statistical Reporting Service Staff Report, Washington, DC, USDA.

Fecso, R., Tortora, R.D. & Vogel, F.A. 1986. *Sampling frames for agriculture in the United States.* Journal of Official Statistics, 2 (3).

Fernández del Pozo, J.M. 1991. *The Spanish programme for the renewal of agricultural statistics.* Proceedings of the Conference on the Application of Remote Sensing to Agricultural Statistics, Belgirate, Italy.

Ford, B.L., Nealon, J.P. & Tortora, R.D. 1986. *Area frame estimators in agricultural surveys: sampling versus non-sampling errors.* Ag. Econ. Res., 38 (2). Journal of the USDA, Washington, DC.

Fournier, P. 1972. *"Etude sur l'utilisation du territoire - Méthodologie ET résultats 1969-1970-71"*. Supplément "Série Etudes", No. 104 (November 1972). 112 p.

Fournier, P. 1986. *"Enquête sur l'utilisation du territoire effectuée par la méthode des segments"*. Paris. Service central des enquêtes et des études statistiques (SCEES).

Fuentes, M. & Gallego, F.J. 1994.*Cluster area estimation on a sample by square grid. Getting a stable variance for the survey of the Spanish Ministry of Agriculture,* p. 146-149. Conference on the MARS Project: overview and perspectives. Luxembourg, Office for Publications of the European Communities.

Gallego, F.J. 1995. *Sampling frames of square segments.* Report EUR 16317. Luxembourg, Office for Publications of the European Communities. ISBN 92-827-5106-6. 68 pp..

Gallego, F.J. & Delincé, J. 1991. Crop area estimation through area frame sampling and remote sensing. In Euro-courses: remote sensing applied to agricultural statistics. JRC. Commission of the European Communities. RSA/91/3.

Gallego, F.J., Delincé J. & Carfagna, E. (1994). *Two stage area frame on squared segments for farm surveys.* Survey Methodology, 20 (2): 107-115.

Gálvez Calix, G. 1982. *El marco muestral de áreas en Honduras.* Dirección General de Estadística y Censos, Honduras.

Garibay, R., Steiner, M. & González, M. 1996. *Area frame point sampling: an exploratory study to measure Nicaraguan agriculture production.* NASS Staff Report. IPO No. 96-01, Washington, DC, USDA.

Geuder, J. 1983. *An evaluation of SRS area sampling frame designs. Ordering count units and creating paper strata.* Statistical Reporting Service Staff Report. Washington, DC, USDA.

Geuder, J. 1984. *Paper stratification in SRS area sampling frames.* Statistical Research Division, Report No. 79. Washington, DC, Statistical Reporting Service. USDA.

Giovacchini, A. 1992. *Agricultural statistics by remote sensing in Italy: An ultimate cost analysis.* Proceedings of the Conference on the Application of Remote Sensing to Agricultural Statistics, Belgirate, Italy.

González, F., López, S. & Cuevas, J.M. 1990. *Estimación de superficies cultivadas en la Comunidad Autónoma de Navarra a partir de imágenes de satélite y encuestas de campo.* Estadíst. Española, 32 (505): 18.

González-Villalobos, A., Moreira, M., Silva, G. & Biffi, A.J. 1988. *Desenvolvimento do sistema de informações agropecuárias.* Annals of the Brazilian Symposium on Remote Sensing. National Space Research Institute, Publ. 4940. Natal, Brazil.

González-Villalobos, A. & Wallace, M.A. 1996. *Multiple Frame Agricultural Surveys - current surveys based on area and list sampling methods.* Vol. I. ISSN 1014-3378. English version. FAO Statistical Development Series No. 7, Rome, FAO. Published in French and Spanish versions, 1997. (Arabic and Chinese translations will be published in 1998.)

Guevara Salinas, R. & Gálvez Cálix, G. 1994. *"Honduras: sistema de encuestas agrícolas utilizando marcos múltiples".* National case study prepared for FAO.

Gurney, M. & González, M.E. 1972. *Estimates from samples from frames where some units have multiple listings.* Proceedings of the Montreal Meetings of the American Statistical Association.

Hanuschak, G.A., Allen, R.D. & Wigton, W.H. 1982. *Integration of Landsat data into the Crop Estimation Program of USDA's Statistical Reporting Service.* 1982 Machine Processing of Remotely Sensed Data Symposium, Purdue University, Indiana, USA.

Hanuschak, G.A. & Craig, M.E. 1993. *Remote sensing program of the National Agricultural Statistics Service: from a management perspective.* Proceedings of the International Conference of Establishment Surveys, Buffalo, NY, USA, 27-30 June, p. 158-166.

Hanuschak, G.A. & Morrisey, K.M. 1977. *Pilot Study of the potential contributions of Landsat data in the construction of area sampling frames.* Statistical Reporting Service Staff Report. Washington, DC, USDA.

Hartley, H.O. 1962. *Multiple frame surveys.* ASA Pro. So. St., 203-206.

Hartley, H.O. 1966. *Systematic sampling with unequal probabilities and without replacement.* Am. Stat. Assn J., 61 (315): 739-748.

Hartley, H.O. 1974. *Multiple frame methodology and selected applications.* Sankhya, Series C, 36, p. 99-118.

Hendricks, W.A., Searls, D.T. & Horvitz, D.C. 1965. *Estimation of areas in agricultural statistics.* Chapter 11. Rome, FAO.

Hill, G. & Rockwell, D. 1977. *Associating a reporting unit with a list frame sampling unit in multiple frame sampling - Ohio and Wisconsin.* Washington, DC, Statistical Reporting Service, USDA.

Houseman, E.E. 1975. *Area frame sampling in agriculture.* Statistical Reporting Service, Report No. 20, English version, Washington, DC, USDA. Also published in a Spanish version, 1976.

Houseman, E.E. & Trelogan, H.C. 1967. *Progress toward optimizing agricultural area sampling.* Proceedings of the 36th Session of the International Statistical Institute, Sydney, Australia.

Huddleston, H.F. 1976. *A training course in sampling concepts for agricultural surveys.* Statistical Reporting Service, Report No. 21 (English version). Washington, DC, USDA. Also published in a Spanish version, 1980.

Ingram, S. & Davidson, G. 1983. *Methods used in designing the National Farm Survey.* American Statistical Association, Proceedings of the Section on Survey Research Methods, p. 220-225.

Jessen, R.J. 1942. *Statistical investigation of a sample survey for obtaining farm facts*. Res. Bull. Iowa St. Univ., 304: 1-104. (Ames, Iowa.)

Jessen, R.J. & Houseman, E.E. 1944. *Statistical investigation of farm sample surveys taken in Iowa, Florida and California for obtaining farm facts*. Res. Bull. Iowa St. Univ., 329: 263-338. (Ames, Iowa.)

Julien, C. & Maranda, F. 1990. *Sample design of the 1988 National Farm Survey*. Survey Methodology, 16 (1): 117-129.

Kibler, W.E. 1966. *Sample survey methods*. Washington, DC, Statistical Reporting Service, USDA.

King, A.J. & Jessen, R.J. 1945. *Master sample of agriculture*. JASA, 40: 30-56.

Kish, L. 1989. *Sampling methods for agricultural surveys*. FAO Statistical Development Series No. 3, Rome, FAO.

Kott, P. S. 1988. *Estimating variances for the June Enumerative Survey*. RAD Staff Report No. SRB-NERS-8806. Washington, DC, NASS/USDA.

Kott, P.S. 1989. *The variance for direct expansions from a common area sample design*. J. Official Statistics, 5: 183-186.

Kott, P.S. 1990. *Mathematical formulae for the 1989 survey processing system (SPS) summary*. Washington, DC, NASS/USDA.

Kott, P.S. & Vogel, F.A. 1995. *Multiple-frame business surveys*. Chapter 11, Business survey methods, p. 185-203. ISBN 0-471-59852-6. New York.

Mahalanobis, P.C. 1944. *On large-scale sample surveys*. Philosophical transactions of the Royal Society of London. Series B, 584 (231): 329-451.

Mayda, J.E. & Chinnappa, N. 1994. *A multiple frame agricultural survey design: a case study from Statistics Canada*. National case study prepared for FAO.

Monroe, J. & Finker, A.L. 1959. *Handbook of area sampling*. Chilton Co.

Moore, R.P. 1969. *Research report on four state multiple frame study*. Statistical Reporting Service Staff Report. Washington, DC, USDA.

Moore, R.P. 1989. *Final report on evaluation of the IBGE Agricultural Harvest Forecasting and Monitoring Project*. UNDP Project. Brazil.

Mueller, C., Silva, G. & González-Villalobos, A. 1988. *Pesquisa agropecuária do Paraná - Safra 1986/87*. ISSN 0034-7175. R. Brasil. Estat., 49 (191): 55-84. (January-June 1988)

Nealon, J. 1981. A*n evaluation of alternative weights for a weighted estimator*. Statistical Reporting Service Staff Report. Washington, DC, USDA.

Nealon, J. 1982. *Response errors in the weighted estimator*. Statistical Reporting Service Staff Report. Washington, DC, USDA.

Nealon, J. 1984. *Review of the multiple frame estimators*. Statistical Reporting Service, Report No. 80. Washington, DC, USDA.

Palacio-Rabaud, V. & Poiret, M. 1996. *"Données chiffrées. Agriculture N° 87-décembre 1996. L'utilisation du territoire en 1996. Nouvelle série de 1991 à 1996"*. Paris, Agreste - La statistique agricole, Ministry of Agriculture and Fisheries.

Porchiér, J.C. 1993. *Ter-Uti surveys, Area estimation by point sampling*. Euro Courses- Remote Sensing Applied to Agricultrural Statistics-Regional Inventories. Athens. (Unpublished)

Pratt, W.L. 1974. *The use of interpenetrating sampling in area frames*. Statistical Reporting Service Staff Report No. 80, Washington, DC, USDA.

Serrurier, D. and Phillips, J.E. 1976. *Double Frame Ontario Pilot Hog Surveys*. Survey Methodology, 2: 138-170.

Silva, G. & González-Villalobos, A. 1987. *Brazil's current agricultural survey*. Contributed paper. 46th session of the International Statistical Institute, Tokyo, Japan.

Silva, G., Martins Carrilho, M. & Nobrega Pitaluga, E. 1994. *Multiple frame agricultural aurveys in Brazil*. National case study prepared for FAO.

Sukhatme, P.V & Sukhatme, B.V 1969. *On some methodological problems of sample surveys of agriculture in developing countries*. Report on the 1960 World Agricultural Census - Methodology, Chapter VII, p. 151-169. Rome, FAO.

Théberge, A. 1994. *A comparison between the Canadian Area Farm Survey and the Census of Agriculture data*. Proceedings of the Annual Research Conference of the United States Bureau of the Census, Washington, DC.

Théberge, A. & Kovar, J.G. 1993. *The design of the Canadian Area Farm Survey*. Presented at the 49th session of the International Statistical Institute, Florence, Italy.

Trépanier, J. & Théberge, A. 1993. *Redesign of the Canadian agriculture commodity surveys*. Proceedings of the International Conference of Establishment Surveys, Buffalo, NY, USA, 27-30 June, p. 688-692.

Turner, C.L. 1993. *A comparison of four alternative weighted estimators to the open estimator for use in the Agricultural Labour Survey*. Proceedings of the International Conference of Establishment Surveys, Buffalo, NY, USA, 27-30 June, p. 818-822.

Vogel, F.A. 1975. *Surveys with overlapping frames - problems and applications*. ASA Pro. So. St., 694-699.

Vogel, F.A. *Survey design and estimation for agricultural sample surveys*. Washington, DC, Statistical Reporting Service, USDA.

Vogel, F.A. & Bosecker, R. 1974. *Multiple frame livestock surveys: a comparison of area and list sampling*. Washington, DC, Statistical Reporting Service, USDA.

Vogel, F.A., Bosecker, R. & Rockwell, R. 1976. *Multiple frame livestock surveys: an evaluation of alternative methods of overlap determination*. Statistical Reporting Service Staff Report. Washington, DC, USDA .

Wallace, M.A. & Alvarado, R. 1982. *Area frame probability samples: design, construction and use of area frames for survey sampling*. Costa Rica, Instituto Interamericano de Cooperación para la Agricultura (IICA). Proyecto de Información Agropecuaria del Istmo Centroamericano (PIADIC). Documentación e Información Agrícola No. 112 (English version). ISSN-0301-438X. (Also published in a Spanish version.)

Wigton, W.H. 1994. *Albania: list/area multiple frame agricultural survey*. National case study prepared for FAO.

Wigton, W. & Bormann, P. 1978. *A guide to area sampling frame construction utilizing satellite imagery*. Statistical Reporting Service Staff Report. Washington, DC, USDA.

Zarkovich , S.S. 1965. *Sampling methods and censuses*. Rome, FAO.

III. OTHER RELATED REFERENCES

ARGENTINA

National Institute of Statistics and Censuses (INDEC). 1996. *Statistical Yearbook of the Argentine Republic, Vol. 12. Encuesta agropecuaria 1992 (datos definitivos)*. Buenos Aires, Ministerio de Economía y Obras y Servicios Públicos.

National Institute of Statistics and Censuses (INDEC). 1996. *Encuesta nacional agropecuaria 1995 - Resultados por zonas estadísticas, Vol. 7.* Buenos Aires, Ministerio de Economía y Obras y Servicios Públicos; Buenos Aires.

Secretary of Agriculture, Livestock, Fisheries and Food. 1993. *Encuesta agropecuaria 1992 (datos definitivos)*. Buenos Aires, Secretaría de Agricultura, Ganadería y Pesca. Ministerio de Economía y Obras y Servicios Públicos; Buenos Aires.

BRAZIL

Brazilian Institute of Geography and Statistics (IBGE). 1994. *Multiple frame agriculture survey in Brazil: interviewer's manual*. English translation prepared for FAO.

COLOMBIA

Departamento Administrativo Nacional de Estadística (DANE). 1994. *El Tamaño de la muestra agrícola de areas*. Report prepared by L.C. Gómez . Santafé de Bogotá, Colombia.

Departamento Administrativo Nacional de Estadística (DANE). 1994. *DANE-Proyecto Sistema de Información Estadístico Colombiano (SISAC). Módulo area, producción y rendimiento.* Procesamiento de la información. Report prepared by L.E. Díaz. Santafé de Bogotá, Colombia.

Departamento Administrativo Nacional de Estadística (DANE). 1996. *DANE-Proyecto Sistema de Información Estadístico Colombiano (SISAC). Módulo area, producción y rendimiento. Manual de organización, logística y supervisión* and *Manual de uso de fotografía aerea y cartografía.* Santafé de Bogotá, Colombia.

Departamento Administrativo Nacional de Estadística (DANE). 1996. *Encuesta nacional agropecuaria: resultados 1995.* Santafé de Bogotá, Colombia.

Departamento Administrativo Nacional de Estadística (DANE). 1996. *Diseño e implementación de la muestra de areas. Documento Técnico No. 3.* Santafé de Bogotá, Colombia.

COSTA RICA

Dirección General de Estadística y Censos. 1989. *Sexta Encuesta Agrícola Nacional. Mayo-junio 1989 - Instructivo del entrevistador.* San José, Ministerio de Economía, Industria y Comercio.

Dirección General de Estadística y Censos. 1993. *Uso del muestreo del areas para las encuestas agropecuarias*. Report prepared by L. Villalobos Rodríguez and P. Solano Mayorga. I Congreso Nacional de Estadística. San José, Dirección General de Estadística y Censos.

FAO. 1994. *Colombia: agricultural area sample surveys in Colombia.* Report prepared for FAO by E. Albarracín Durán.

FAO. 1994. *Egypt: Area frame survey in Egypt.* Report prepared for FAO by M.S.M. Nazif.

FAO. 1995. *Programme for the World Census of Agriculture 2000.* FAO Statistical Development Series No. 5, Rome, FAO. Also published also in Arabic, French and Spanish versions, 1996.

FAO. 1995. *Conducting agricultural censuses and surveys*. FAO Statistical Development Series No. 6, Rome, FAO. Also published also in French and Spanish versions, 1996.

FRANCE

Ministry of Agriculture and Fisheries. 1994. *Enquête: Utilisation du territoire. Instructions aux enquêteurs,* Paris. Service central des enquêtes et des études statistiques (SCEES) (April, 1994).

IFDC/AAIC. 1992. *Summary report. Area sampling frame survey in Albania.* Muscle Shoals, Ala., USA, Agricultural Assessments International Corporation (AAIC)/International Fertilizer Development Center (IFDC).

ITALY

Ministry of Food, Agriculture and Forestry. 1993. *AGRIT 93, Innovazione tecnologica. Sistema statistico provinciale.* Edited by Consorzio Italiano per il Telerilevamento in Agricoltura, Rome.

Ministry of Food, Agriculture and Forestry. 1995. *AGRIT 11, Specifiche tecniche per le rilevazioni.* Edited by Consorzio Italiano per il Telerilevamento in Agricoltura, Rome.

Ministry of Food, Agriculture and Forestry. 1995. *AGRIT 11, Sperimentazione nuove procedure statistiche, Campania, Marche, Veneto.* Edited by Consorzio Italiano per il Telerilevamento in Agricoltura, Rome.

Ministry of Food, Agriculture and Forestry. 1996. *Remote sensing in agriculture. Crop acreage estimate and crop production forecasts.* Edited by Consorzio Italiano per il Telerilevamento in Agricoltura, Rome.

Ministry of Food, Agriculture and Forestry. 1996. *AGRIT 12, Statistiche agricole anno 1996.* Bulletin F1, 31 December. Edited by Consorzio Italiano per il Telerilevamento in Agricoltura, Rome.

MOROCCO

Ministry of Agriculture and USAID. 1991. *Présentation du système aréolaire.* Ministère de l'Agriculture et de la mise en valeur agricole, Rabat.

SPAIN

Ministry of Agriculture, Fisheries and Food. 1990. "*El marco de areas como instrumento de base para la estadística de superficies de cultivo*". Bol. Mens. Estadíst., 12: 85-106.

Ministry of Agriculture, Fisheries and Food. 1991. *"Encuesta sobre superficies y rendimientos en Castilla y León, en 1990"*. Bol. Mens. de Estadíst. 6 (June).

Ministry of Agriculture, Fisheries and Food. 1992a. *"Utilización de imágenes de satélite para la estimación de superficie de cultivos en la provincia de Segovia, en 1990"*. Bol. Mens. de Estadíst. (March).

Ministry of Agriculture, Fisheries and Food. 1992b. *"Aplicación de la teledetección en la provincia de Cuenca"*. Bol. Mens. de Estadíst. (May).

Ministry of Agriculture, Fisheries and Food. 1992c. *"Encuesta sobre superficies y rendimientos de cultivos del año 1990"*. Bol. Mens. de Estadíst. (April).

UNITED STATES

USDA. 1968. *Report on 1962-1963 Research Projects.* Statistical Reporting Service Staff Report. Washington, DC, United States Department of Agriculture.

USDA. 1983. *Scope and methods of the Statistical Reporting Service.* Miscellaneous Publication No. 1308. Washington, DC, United States Department of Agriculture.

USDA. 1984. *Area frame analysis package.* Statistical Reporting Service Staff Report. Washington, DC, United States Department of Agriculture.

USDA. 1987. *Basic Agricultural Survey Statistics and Methods.* A training course prepared by the National Agricultural Statistics Service. Washington, DC, United States Department of Agriculture.

USDA. 1992. *Supervising and editing manual - agricultural surveys*. Washington, DC, National Agricultural Statistics Service, United States Department of Agriculture.

USDA. 1994. *CASS. Computer-aided stratification and sampling*. Washington, DC, National Agricultural Statistics Service, United States Department of Agriculture.

USDA. 1996a. *Preparing statistics for agriculture*. Washington, DC, National Agricultural Statistics Service, United States Department of Agriculture.

USDA. 1996b. *Agricultural surveys. Interviewer's manual. June 1994-May 1995*. Washington, DC, National Agricultural Statistics Service, United States Department of Agriculture.

USDA. 1997. *June and fall area frame design information for 1997*. Washington, DC, National Agricultural Statistics Service, United States Department of Agriculture.

National agricultural survey programmes, established to obtain reliable and timely basic data on the agricultural sector, are based on one of three sampling survey statistical methods: list sample designs, area sample designs, and multiple frame designs, in which a dual sampling frame is utilized, composed of a combination of an area frame and a list frame. *Multiple frame agricultural surveys* is a two-volume comprehensive introduction to establishing and conducting area and multiple frame probability sample survey programmes, emphasizing methods and practices applicable in developing countries. The first volume, published in 1996 (FAO Statistical Development Series No. 7), provides a general classification of alternative agricultural survey designs, with an indication of their respective advantages and limitations. It examines the points to consider in establishing and conducting a periodic agricultural survey programme based on dual frame sampling methods, i.e. probability selection and estimation methods, the survey organization, the equipment and materials needed, data collection, summarization and processing. It also includes a detailed description of a category of survey designs considered especially useful for developing countries. This second volume presents the survey methodologies of area and multiple frame agricultural survey programmes currently used in a wide range of countries. It provides actual examples of the application of the survey methods presented in the first volume. The purpose is to allow the reader to make the necessary comparisons and to facilitate the application of the most appropriate statistical methods and practices.

ISBN 92-5-104074-5 ISSN 1014-3378

9 789251 040744

M-77 W7334E/1/5.98/1500