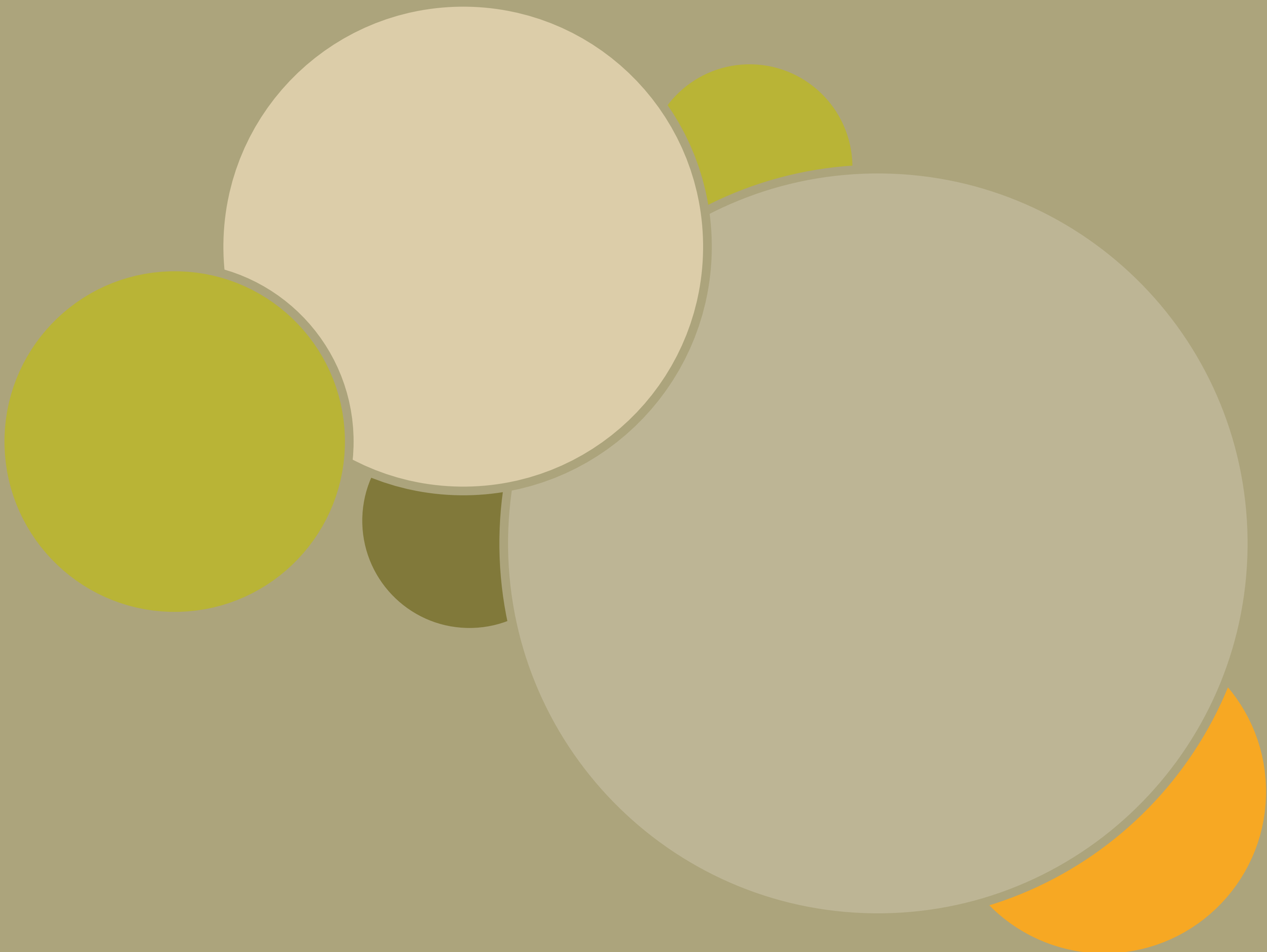
Food and Agriculture Organization
of the United Nations

Handbook on crop statistics: improving methods for measuring crop area, production and yield

Publication prepared in the framework of the Global Strategy to improve Agricultural and Rural Statistics

Handbook on crop statistics: improving methods for measuring crop area, production and yield

December 2018



Contents

Tables, figures and boxes

TABLES

FIGURES

BOXES

Acronyms

ADB	Asian Development Bank
AGRIS	Agricultural Integrated Survey
APS	Agricultural Production Surveys
BOS	Bureau of Statistics
BAS	Bureau of Agricultural Statistics
CA	Census of Agriculture
CAPI	Computer Assisted Personal Interview
CC	Crop Cutting
CCE	Crop Cutting Experiment
CCM	Crop Card Monitor
CES	Crop Estimation Survey
CES-F&V	Crop Estimation Survey of Fruits and Vegetables
CHAMAN	Coordinated Programme on Horticulture Assessment and Management using geoiNformatics
CU	Counting Units
DEFRA	Department of Environment, Food and Rural affairs
DES	Directorate of Economics and Statistics
DESMOA	Directorate of Economics and Statistics, Ministry of Agriculture
EA	Enumeration Area
EPSEM	Equal Probability Selection Method
FAO	Food and Agriculture Organization of the United Nations
GCES	General Crop Estimation Survey
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
GSARS	Global Strategy to improve Agricultural and Rural Statistics
IASRI	Indian Agricultural Statistics Research Institute
IBGE	Instituto Brasileiro de Geografia e Estatística
IDA	International Development Association
ISU-CSSM	Iowa State University – Center for Survey Statistics and Methodology
JSAH	June Survey of Agricultural and Horticultural Activities
LBC	Licensed Buying Companies
LSMS	Living Standard Measurement Survey
LSMS-ISA	Living Standard Measurement Survey – Integrated Surveys on Agriculture
MAFW	Ministry of Agriculture and Farmers Welfare
MCP	Measuring Cassava Productivity
MOFA	Ministry of Food and Agriculture
MSF	Master Sampling Frame
MSS	Multispectral System
NASS	National Agricultural Statistics Service
NDVI	Normalize Difference Vegetation Index
NHB	National Horticultural Board
NRCS	Natural Resources Conservation Service
NRI	National Resources Inventory
NSO	National Statistical Office
PCPS	Palay and Corn Production Survey

PLLS	Public Land Survey System
PPP	Plant protection Product
PSU	Primary Sampling Unit
RCPS	Rice and Corn Production Survey
SAR	Satellite Aperture Data
SLR	Standard Labour Requirements
SNPA	Sistema Nacional de Pasquisas por Amostragem de Esabelecimentos Agropecuarios
SRS	Simple Random Sampling Without Replacement
SRID	Statistics, Research and Information Directorate
SSU	Secondary Sampling Unit
TSL	Taylor Series Linearization
UNPS	Uganda National Panel Survey
USDA	United States Department of Agriculture
USU	Ultimate Sampling Unit
UT	Union territory
WCA	World Census of Agriculture
WFA	Whole Farm Approach

Acknowledgments

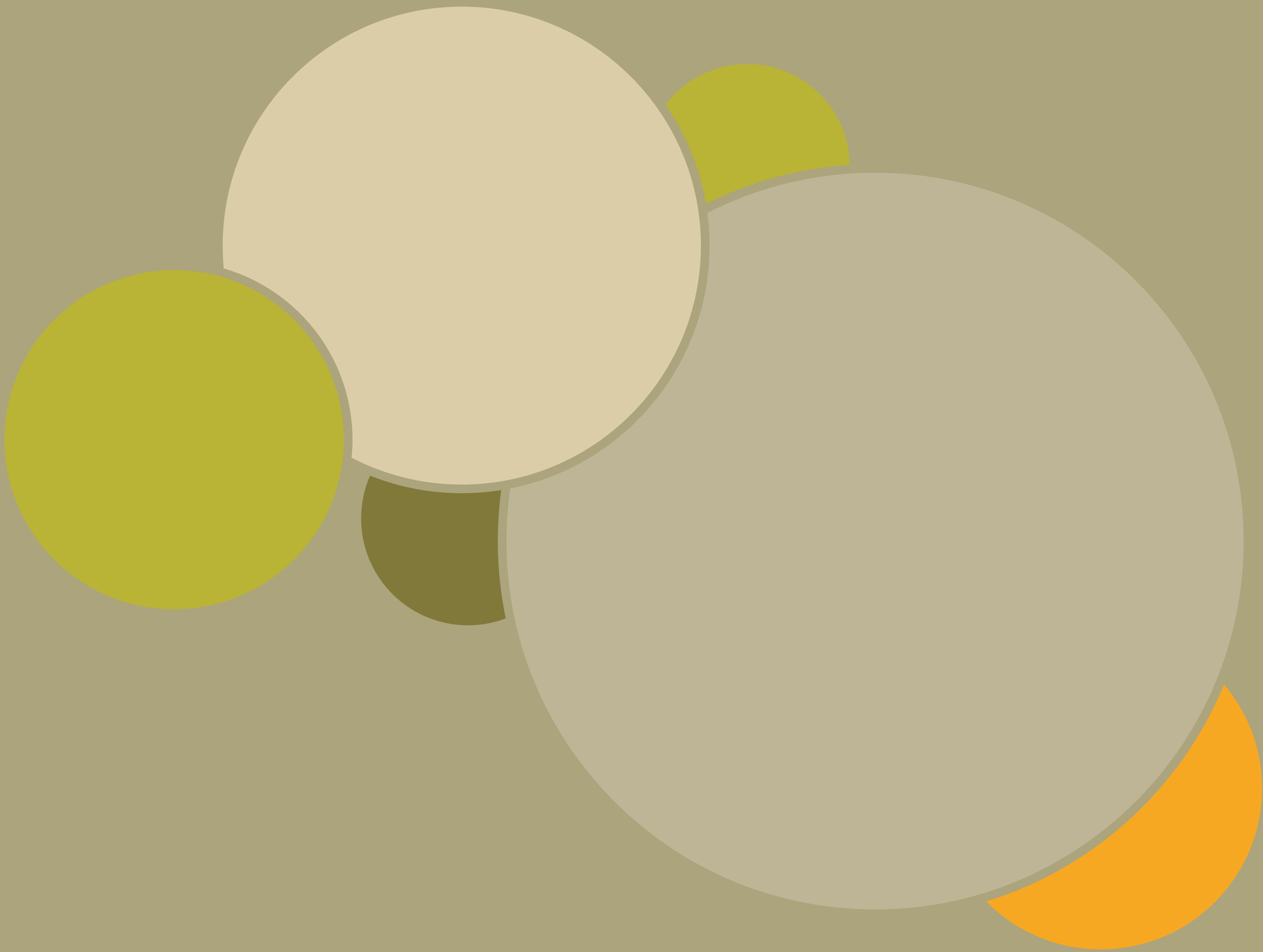
This Handbook is the result of a research project undertaken within the Global Strategy to improve Agricultural and Rural Statistics (GSARS), a capacity-building initiative in the field of agricultural statistics hosted by the Statistics Division of FAO.

The main author of the Handbook is Dalisay S. Maligalig. She was assisted by George Mantey Asamoah, Clifford B. Lesmoras, Hazel M. Remolacio and Charissa N. Castillo. Heather Moylan and Robert Arcaraz wrote chapters 7 and 8, respectively. Valuable inputs were provided by Naman Keita, Flavio Bolliger, Neli Georgieva Mihaylova, Nayo Ankouvi, Talip Kilic, Michael Austin Rahija, Valérie Bizier, and Arbab Asfandiyar Khan. Clara Aida Khalil acted as focal point of this research project, both coordinating and participating in the technical developments.

The major results used in drafting the Handbook were reviewed during the Fifth Scientific Advisory Committee Meeting and an Experts' Meeting held at FAO Headquarters on 7–8 September 2016. Relevant comments made during these events were taken into account in drafting the Handbook.

This publication was prepared with the support of the Trust Fund of the Global Strategy, funded by the United Kingdom's Department for International Development (DfID) and the Bill & Melinda Gates Foundation. The World Bank also provided financial and technical support towards drafting the publication.

Arianna Martella coordinated the design and communication aspects. The document was edited by Sarah Pasetto and formatted by Laura Monopoli.



Preface

Agriculture is still the most important sector of national economies in many developing countries. Agricultural statistics are a key prerequisite for the management of the sector. They provide information required to monitor trends and estimate the future prospects of agricultural commodity markets, and, therefore, to assist in making policies on aspects such as price support, imports and exports, and distribution. In particular, statistics on crop area, production and yield play a central role in the planning and allocating of resources for the development of the agricultural sector.

Many developing and underdeveloped countries, however, still lack the capacity to produce and report even the minimum set of core agricultural data. As a response to this issue, at its Forty-first session in February 2010, and after an extensive consultation process with national and international organizations, the United Nations Statistical Commission (UNSC) endorsed the Global Strategy to improve Agricultural and Rural Statistics (GSARS).

GSARS is a capacity development initiative for improving the availability of data on the agricultural and rural sectors, necessary for evidence-based decision-making. This initiative is implemented through a Global Action Plan, which defines the technical assistance, training and research plans as well as the governance mechanisms. Under the GSARS Research component, one of the priority areas of work identified was the development of Improved Methods for Producing Crop Statistics (namely, statistics on crop area, production and yield). Accordingly, a series of projects under specific domains of crop statistics were conducted during the first phase of GSARS, which resulted in the production of a series of working papers and technical reports to serve countries seeking to improve the quality of their crop statistics. More specifically, projects aiming to bring about improvements in the following domains were undertaken:

- estimation of area, production and yield for mixed, repeated, and continuous cropping;
- estimation of area, production and yield of root crops;
- estimation of area, production and yield of vegetable crops.

This publication is, on one hand, an attempt to pull all these efforts together into a comprehensive Handbook on methods for producing crop statistics. On the other, it updates a previous publication of the Food and Agriculture Organization of the United Nations (FAO), entitled Estimation of crop areas and yields in agricultural statistics, which – despite being a reference point in the field of crop statistics – dates back to 1982.

1

Introduction

This Handbook aims to serve as a guide for statisticians from National Statistical Offices (NSO) seeking to improve their data collection methods to produce crop statistics, and to become the new reference document for estimating crop area, yield and production.

The present Handbook updates and upgrades a previous publication of the Food and Agriculture Organization of the United Nations (FAO), entitled *Estimation of crop areas and yields in agricultural statistics*, which – despite being a reference point in the field of crop statistics – dates back to 1982. This Handbook takes into account new methodological developments and technological advances made since then. It builds upon the results of methodological studies conducted in the context of the Research component of the Global Strategy to improve Agricultural and Rural Statistics (GSARS or Global Strategy).

More specifically, the following GSARS publications have served as key references for this Handbook:

- Methodology for Estimation of Crop area and Crop Yield under Mixed and Continuous Cropping (Sud, U.C. *et al*., 2017).
- Measuring Crop Area and Yield under Pure Stand, Mixed and Continuous Cropping: Findings from the Field Tests in three countries (Sud, U.C. *et al*., 2016).
- Crop Yield Forecasting: Methodological and Institutional Aspects (Bernardi, M. *et al*, Delincé J., Durand W., and Zhang N., 2016).
- Synthesis of Literature and Framework – Research on Improving Methods for Estimating Crop Area, Yield and Production under Mixed, Repeated and Continuous Cropping. (Sud, U.C. *et al*., 2016).

Furthermore, the Handbook largely draws upon FAO publications on measuring crop area and yield and multiple frame surveys, and from the World Bank's Handbook on Land Area Measurement (e.g. Carletto, G. *et al*., 2016a).

The experience of some developed countries that have upgraded their data collection systems with technologies that help to obtain accurate and timely crop statistics have been also taken into account. In this respect, the Handbook highlights the lessons derived from these methodological studies, to enable statisticians in developing countries to adapt them for use in their own systems.

Considering that resources allocated to agricultural statistics and the skill level of staff may vary across countries, the Handbook recognizes that a single data collection method addressing all the challenges associated to crop statistics cannot be identified. Therefore, the publication extensively describes each data collection method and guides the reader through their respective strengths and weaknesses.

Chapter 2 presents the alternative uses of crop statistics, which include policy formulation and monitoring, in designing and tracking interventions by various government agencies. The characteristics, strengths and weaknesses of data sources used in compiling crop statistics are fully examined. Furthermore, the main challenges arising when striving to obtain good-quality data are presented. Of the data sources discussed in chapter 2, probability sample surveys are the most reliable source of crop statistics between agricultural censuses. Approaches to designing and estimating probability agricultural sample surveys are discussed in chapter 3, taking into account other relevant Global Strategy publications, particularly the Handbook on Master Sampling Frames for Agricultural Statistics: Frame Development, Sample Design and Estimation and the Handbook on Remote Sensing for Agricultural Statistics. These approaches vary according to the capacity of countries to construct sampling frames and the available resources for conducting nationwide surveys. Relevant aspects of designing a probability sample survey, including definition of the survey objectives, determination of the sample size and sample allocation, appropriate sampling techniques, definition of survey weights, and estimation procedures are presented in the chapter. The processes to be followed in constructing a list frame, an area frame and multiple frames, as well as the corresponding selection and estimation procedures, are also discussed.

The most important data in a crop production probability sample survey are the area, the production and the yield of specific crops. The two methods for collecting these data are farmer declarations or interviews and objective measurements. When data are gathered by interviewing holders or farmers, they is asked to report their total production and area planted and harvested. The yield is then computed as the ratio of total production and harvested area. On the other hand, when objective measurement is employed, the area planted is measured and the average yield is derived through crop-cutting experiments. These two methods, as well as a number of alternative methods, are discussed in chapters 4 and 5. Specifically, chapter 4 explores various methods for measuring crop area, including the use of remote sensing, while chapter 5 discusses procedures for measuring crop production or yield.

Smallholders in developing countries form the majority of the rural population. To maximize the area that they are able to cultivate, farmers often practice mixed or continuous cropping. Obtaining reliable data on crop area in mixed fields can be challenging, especially when several crops are cultivated in the same field. Research has shown that when mixed crops are not properly estimated and recorded, productivity and consequently income will be underestimated. Appropriate methodologies for better estimation of crop area and strategies for estimating crop production under mixed cropping are examined in chapter 6.

Chapter 7 presents approaches to obtaining reliable production data on crops that have staggered maturity and need to be harvested several times. The chapter discusses the challenges to be faced in, and procedures for, measuring the production and yield of root and tuber crops, which are good examples of crops with staggered maturity (requiring continuous harvest). Based on the results of two randomized household survey experiments conducted by the World Bank on cassava, an important root crop in many developing countries, chapter 7 provides recommendations on how to improve data collection for this kind of crop.

Vegetable crops represent the most important component of horticulture, which is a continuously expanding sector in many developing countries. Despite the social and economic importance of cultivating vegetable crops, information on this sector remains scarce, also because of the objective difficulties in measuring the area and production of this crop category. Chapter 8 discusses the main issues related to the production of statistics on vegetable crops, presenting the various methodological options available. Finally, chapter 9 concludes this publication with a discussion of the use of this Handbook in integrated agricultural survey programmes.

2

Relevance and use of statistics on crop area and production

The economies of most developing countries are often characterized by high population growth rates and low farm productivity that, in turn, could result in food shortages if appropriate policies are not put in place. To mitigate this problem, the strategic development objectives of these countries must include ensuring food security, increasing farm incomes and reducing income inequalities by promoting the development of the agricultural sector. However, formulating and monitoring the implementation of effective agricultural policies requires timely and reliable agricultural data, including crop production statistics, which are still rarely produced in most developing countries.

This chapter illustrates the main uses of statistics related to crop area and production, the available data sources, and the challenges to address when seeking to achieve good quality data.

2.1 USES OF CROP PRODUCTION STATISTICS

Crop production statistics support policy formulation and monitoring in the fields of health, food security, natural resource use, renewable energy production, environmental stewardship, rural development, crisis management during natural calamities, and assessment of the long-term viability and competitiveness of the food chain. Crop statistics are used by various government institutions and other sectors of society for planning, policy-making and benchmarking.

Agricultural operators who make decisions on what crops to plant, when to plant them, what seed varieties to use, and what cultural practices to implement could benefit from these statistics. Farmers in developed countries export their produce and require access to the production data relating to their neighbouring areas and other countries for guidance in making business decisions. Farmers in developing countries, however, are more concerned with obtaining sufficient produce for their family's subsistence; still, they also need data to decide how much area to cultivate, what crops to plant and when to plant them.

Businesses also use crop production data, just as farmers, food wholesalers, retailers and traders need information on crop production at subnational/local levels. Agricultural suppliers also employ crop production data to plan marketing strategies.

While farm-level decisions are made by farmers, the government can create the enabling environment so that its development agenda can be achieved (Colwell, 1989). To do so, the government needs good-quality data to support decisions and for the purposes of economic planning and research. For example, ministries of agriculture may need time series of production estimates to assess the impact of a specific programme, such as maintaining an adequate supply of food for the population (the most critical problem in developing countries). In many countries in Asia where rice is the staple food, it is important to monitor rice production so that when production and inventory are insufficient, an adequate amount can be imported to cover the shortfall. In this case, the trends of rice production and area planted and harvested must be examined. Data on relevant inputs such as fertilizers, seed varieties and the use of irrigation must also be studied. A spatial analysis of the patterns of consumption and production may also be done if subnational data are available. These analyses will form the basis of government decisions when setting up priorities and policies to maintain food security.

The agencies that provide extension and other services (such as subsidies) to agricultural communities need data to measure the efficiency and productivity of agricultural outputs. Government agencies that deliver subsidies and food aid need data to assess food security at subnational level. During emergencies and disasters, land use and crop production statistics aggregated at subnational levels are used to estimate crop damages, identify the most vulnerable areas that need urgent help and plan the recovery phase. These data are also used by governments or private companies that provide crop insurance to farmers. Subsidies and other resources are also allocated by governments based on agricultural data, including crop area and production.

In general, good-quality crop production statistics are essential in crafting and monitoring programmes and policies aimed at enabling countries to achieve food security and economic progress. These estimates are also used for planning by both the government and the private sector and for resource allocation, as they enable identifying areas at risk and developing good intervention approaches. The international development community, including intergovernmental organizations, also use crop statistics for policy formulation and monitoring, as well as for resource allocation. Finally, statisticians need crop production statistics to compile the gross value added of agriculture and other national accounts statistics.

FIGURE 2.1. STAKEHOLDERS NEEDING GOOD-QUALITY DATA ON CROP AREA AND PRODUCTION.

Source: Author elaboration, 2018.

All these functions require that crop production estimates be precise, timely and regularly updated. Inaccurate data may lead to inappropriate policies and misallocation of the limited resources available. For example, if rice production is overestimated and inventory stocks and imports in the pipeline are not enough to cover the gap, then a shortage of the staple food would arise. Decisions on the volume of commodities to import or export depend on the available production and consumption data. Similarly, if rice production is underestimated and the import pipeline and inventory are rather large, then the surplus rice may spoil, causing the country to make unnecessary expenditures that could have been allocated to other important needs. Therefore, it is paramount to maintain a high level of data quality.

2.2 DATA SOURCES

Crop production estimates are the product of two components: crop area and yield per unit of area. Precise estimations of both harvested area and yield are equally important in ensuring the accuracy of their product. Data on area harvested, yield or production can be collected through agricultural censuses, crop production surveys, other household surveys or administrative data.

Agricultural censuses are periodically conducted by many countries to collect data on the structure of agriculture, in line with FAO recommendations regarding its decennial programmes. Data from these censuses are also used by NSOs to construct sampling frames for agricultural surveys. Agricultural censuses are implemented every five or ten years, during which a complete enumeration of farm holdings is usually done. However, some countries with limited resources, such as Nepal or the Philippines in 2001, used a large sample or a sample census in lieu of complete enumeration. The coverage of agricultural censuses varies across countries. A threshold on the area, volume or value of output, number of livestock or trees, or labour requirements may be imposed in the definition of holdings. Definitions of crop area also differ across countries. Commercial farms producing forestry products may also be included by some countries.

In many cases, agricultural censuses only collect structural data such as crop area, while data on production are often not included. On the other hand, some agricultural censuses also collect information on farm equipment and machinery, transport equipment and types of irrigation.

When an agricultural census conducts a complete enumeration of holdings, summary statistics at finer levels of aggregation can be derived. However, as data is collected every five or ten years, the summary statistics become obsolete over time. Also, because of the large volume of data, the processing and validation of census data requires more time compared to survey data. While the sampling error is assumed to be nil in a complete enumeration, non-sampling errors such as enumerator and supervisor errors in recording the data, respondent errors and data processing errors may still occur. These errors, unlike sampling errors, cannot be estimated; however, they are usually controlled through good field operations and data processing practices.

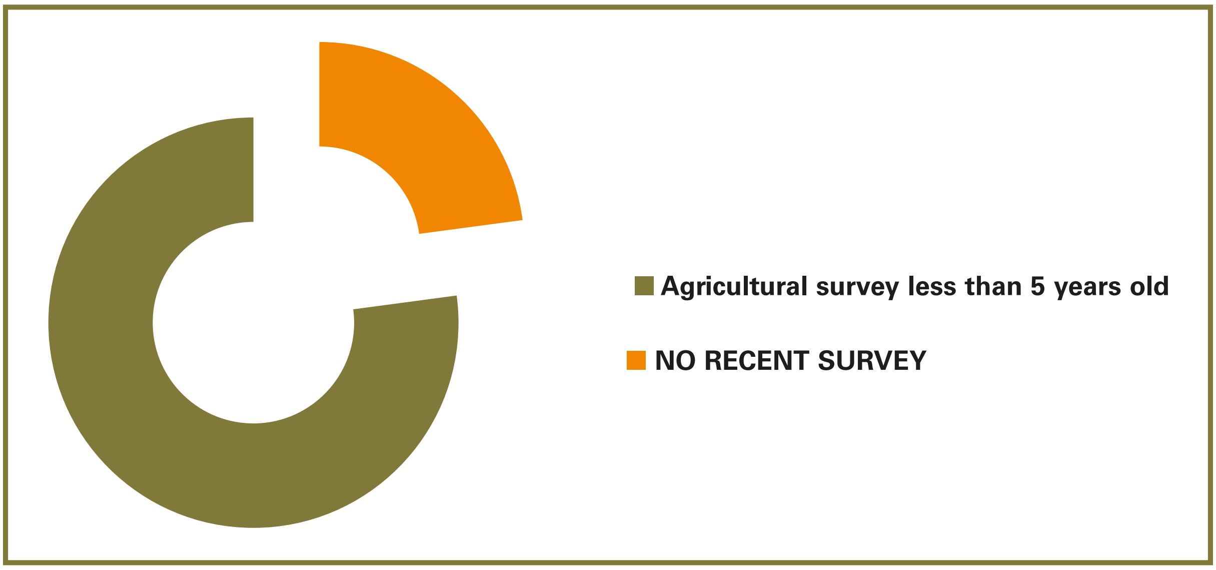
Probability agricultural sample surveys are credible sources of agricultural statistics. They can provide reliable statistics on crop production or yield and area. Their characteristics vary across countries, depending on available resources and requirements. Developed countries typically conduct a specific crop production survey using multiple frames per cropping season, while developing countries usually have one annual general-purpose agricultural survey. Some developing countries may also conduct quarterly production surveys for major crops and livestock.

If data are properly collected using probability sample surveys, the quality of the resulting estimates can be gauged using their corresponding sampling errors. Well-designed probability sample surveys are expected to provide reliable estimates. Because the volume of data from a sample survey is more manageable compared to that of data from an agricultural census, data are processed, analysed and disseminated promptly. While a sample survey is less costly than a census, it still requires substantial human and financial resources to ensure that the estimates are of good quality and that subnational level estimates are obtained with acceptable sampling errors. As the desired level of disaggregation becomes finer, the minimum sample size that will render a tolerable sampling error increases and, therefore, the cost of field operations will also increase. If field operations require objective area and yield measurements, additional surveying costs should be allocated.

Based on the list of core data items established by GSARS and the information gathered from many countries, particularly in Asia and the Pacific, *administrative reporting systems* are among the three major sources of agricultural statistics. Many developing countries that are able to conduct surveys and censuses only when funds from external sources are available rely heavily on administrative reporting systems to compile agricultural production data. Based on the country assessment questionnaire and data and information independently gathered by the World Bank and the Asian Development Bank (ADB) in 2011, of the 57 countries or economies in the Asia and the Pacific region, only 34 have conducted agricultural censuses and 25 hold regular crop production surveys for their major crops.

A similar assessment was performed in the context of GSARS when developing the methodological plan for the Agricultural Integrated Survey (AGRIS) project. To evaluate the availability of agricultural survey data in the 75 countries eligible for International Development Association (IDA) resources,[1] an extensive review of existing data collection programs was performed. As a result, experts from GSARS highlighted that the vast majority of these countries had not conducted agricultural annual surveys in the five years before the assessment (GSARS, 2018).

FIGURE 2.2. THE LACK OF AGRICULTURAL SURVEYS AND CENSUSES IN IDA COUNTRIES.



Source: GSARS, 2018.

Although the country assessment questionnaire used by ADB did not contain direct questions on the use of administrative reports and data, it can be inferred that many countries still rely on this system, which requires a minimal budget from NSOs but is capable of producing statistics on a timely basis. The same conclusion was also drawn for African countries. In their review of assessment reports from developing countries, Keita and Chin (2013) noted that in the absence of a system of agricultural surveys and censes, developing countries use statistics from a variety of sources, including administrative reporting systems and other types of administrative data.

1 http://ida.worldbank.org/about/borrowing-countries

There are several models of administrative reporting systems for agricultural production in developing countries. However, all of these models share the same process: summary statistics collected by the lowest government unit are transmitted to the next step in the hierarchy until they reach the national level. In some countries, government personnel working in the field of agriculture would assess crop production by observing the harvests and, more often, interviewing experts (village heads, farmers, traders, etc.) in their assigned localities as the initial data collection point. In other countries, the village head is the initial data collection point of the process. While this data collection approach is inexpensive and can provide timely data at a finer disaggregation level, research has shown that it is prone to large measurement errors. Those involved in the data collection process have their own vested interests, which could influence the subjective estimation process. Agricultural officers tend to overestimate production in their respective assigned areas to support their claims of accomplishment. In some countries that are transitioning from centrally planned to market-based economies, village heads are likely to report the planned or target figures even when there is a shortfall in the actual production figures. However, as this type of administrative reporting system does not include a validation process that could improve the quality of estimates, the resulting national level estimates are perceived to be biased.

Land registers or land cadastres in which the coverage, value and ownership of real property is registered together with the corresponding map (FAO, 1982) and other administrative data sources, such as specific agencies, boards or organizations established to deal with specific crops may be used to supplement or validate crop production and area estimates. These administrative data could also be used to establish a statistical farm register that can be employed as a sampling frame for a probability agricultural survey or for direct tabulation (Berg and Li, 2015). There are challenges to address in combining administrative data from various sources because the observational units, concepts and definitions of various sources may differ. In developed countries, the data are likely to be digitized and include accompanying metadata; however, in developing countries, the administrative data usually remain in various reporting forms, thus making them suitable only for primary uses.

BOX 2.1. USING AGRIS TO COLLECT DATA ON CROP AREA AND PRODUCTION.

AGRIS was developed in the context of GSARS in response to the need for better, cost-effective and timely statistical data in the agricultural and rural sector. Data generated through this survey programme are intended to inform policy-making and improve market efficiency.

AGRIS is a farm-based modular ten-year survey programme designed as a cost-effective tool to help national statistical agencies accelerate the production of quality disaggregated data on the tecÚical, economic, environmental and social dimensions of farms, including smallholders. It consists of a core module, which should be implemented every year, and a set of four additional rotating modules devoted to specific themes, which are surveyed at a lower frequency. For details on the full methodology proposed to implement AGRIS, the Handbook on the Agricultural Integrated Survey is an exhaustive reference manual (GSARS, 2018).

The AGRIS core module devotes an entire section to crop production activities carried out during the survey reference period. In this section, the respondent, who is normally a member of the holding and is informed on all its activities, is asked to report the area planted, area harvested and quantity harvested of each cultivated plot. Both area and production are surveyed through the farmer declaration approach illustrated in chapters 4 and 5. Additional information that is relevant to crop production and is collected through AGRIS include use of fertilizers, plant protection products and irrigation.

More details on the AGRIS programme are provided in chapter 9.

2.3 MAIN CHALLENGES

To be considered of good quality, data must be relevant, accessible, timely, consistent, clearly interpreted and accurate. For crop production statistics to be accurate, the data collection methods must be designed in such a way as to keep the two components of the total error of an estimate – bias and variance – at tolerable levels. In addition, it is necessary to collect data on a regular basis. Estimates derived from an administrative reporting system are perceived to be affected by large measurement errors, as described above. However, many developing countries continue to use data from the system because they cannot afford to conduct periodic agricultural sample surveys. An administrative reporting system can be improved by incorporating a data validation mechanism in the process, such as an audit sampling, by training data collectors to apply uniform definitions and concepts and by automating the reporting process.

A census with complete enumeration will not incur sampling errors, although it may present bias from measurement and coverage errors. Unless there is a post-enumeration survey, these errors cannot be estimated. However, they could be controlled through effective field operation procedures, well-trained enumerators and supervisors and efficient implementation of a data processing strategy. Moreover, while censuses could provide data at a finer disaggregation level, as they are usually conducted every ten years, crop production statistics tend to become obsolete before updates are available.

Estimates derived from sample censuses and probability surveys may present both sampling and non-sampling errors. Non-sampling errors, including measurement errors, cannot be measured but can be controlled. Sampling error, on the other hand, is a good measure of the reliability of an estimate derived from a probability sample survey. A probability sample survey that is implemented well in the field is considered a better source of crop production statistics compared to an administrative reporting system. Because of the limited resources allocated to agricultural statistics, however, developing countries could conduct only a multipurpose agricultural survey to derive all production estimates of major crops and livestock, instead of the practice of conducting specific crop surveys for every season.

The approach applied to obtain quality estimates using probability sample surveys could also vary across countries because of differences in the characteristics of agricultural farms. Table 2.1. shows the average farm size and number of parcels in countries by continent, based on data from agricultural censuses conducted between 1995 and 2005. While the average farm size of the 114 reporting countries is approximately 5.5 hectares (ha), the range of average farm sizes varies across countries. In the Asia and Pacific region, the average farm size is 1 ha, while it is 117 ha for North American and Central American countries. In the past decade, the average size of holdings has declined in many developing countries in Asia and Africa (Som, 2010). Furthermore, based on the most recent agricultural censuses, small agricultural holdings comprise the majority of farms in many developing countries.

TABLE 2.1. AVERAGE AREA AND NUMBER OF PARCELS PER HOLDING, BY CONTINENT.

Countries by continent (number of reporting countries is given in parenthesis)	Average area per holding (ha)	Average number of parcels per holding
World total (114)	5.5	3.5
Africa (25)	11.5	3.0
America, North & Central (14)	117.8	1.2
America, South (8)	74.4	1.2
Europe (29)	12.4	5.9
Asia (29)	1.0	3.2

Source: Som, 2010.

Obtaining quality estimates from smallholders in developing countries, however, can be challenging. With the decreasing size of holdings and varied tenurial arrangements, identification of the appropriate agricultural operators or farm holders is not straightforward, which can make construction of a good sampling frame rather challenging. Cadastral maps that can be used in developing area frames for probability sample surveys and censuses are often unavailable or not up-to-date. While the costs of relevant technologies (including satellite imagery) have become affordable, often, the staff of many NSOs in developing countries have not yet acquired sufficient skills to improve their sampling frames and crop production estimates.

With the decrease in average farm size, farmers in developing countries have adopted mixed cropping to increase productivity and to balance the soil nutrients of their farms. Some farmers may also diversify to crops with better commercial value, while others may still produce mainly for subsistence purposes and inputs into other household production activities. With the variety in farm practices and production strategies, and considering the food requirements of smallholders' own households, the planting and harvesting of crops could take place all year round. Because of climate change, cropping patterns also vary from season to season. The growth cycles of crops are likely to vary and require more complex field operations if production data is to be accurately recorded. Moreover, farmers whose primary concern is to produce sufficient food for their family's subsistence do not usually document their transactions regarding crop production and household consumption. Therefore, these farmers may not be able to provide reliable estimates of their crop production and area. If resources are available, then objective measurements can be incorporated into survey operations to reduce measurement errors.

Outside the Global Strategy, there have been insufficient studies on how good data collection methods that are already being implemented by countries with advanced statistical systems can be adapted by developing countries that have limited resources to devote to agricultural statistics. Research on the control of non-sampling errors and a comparison of the quality of the data collected using various sources and methods have not been done extensively because NSO staff have limited research capacity. Moreover, most of the resources available to NSOs are appropriated for data collection, leaving very little for analysis and methodological research.

To address this gap, relevant methodological research that has been tested in selected developing countries and data collection methods proven to be effective have been compiled into this Handbook, which is envisaged to be a good reference manual for NSOs seeking to improve their measurement of crop production or yield and area.

3

Sampling and estimation methods for crop statistics

This chapter examines various approaches for designing a probability sample for a crop production survey. The first section discusses the planning activities, consisting of defining the survey objectives, determining the data items that will be included and how they will be collected. The subsequent sections focus on the development of a viable sampling strategy, including the choice of a suitable sampling frame, sample size determination and allocation, selection methods, and estimation. Variance estimation procedures and methods to compensate for noncoverage and nonresponse errors are described. Some sampling strategies and the estimation of probability surveys using three types of sampling frames – list, area, and multiple frames – are also presented.

3.1 ESTABLISHING THE OBJECTIVES OF THE SURVEY, TARGET POPULATION, DATA ITEMS TO BE COLLECTED, FREQUENCY OF COLLECTION

In general, when a sample survey is to be conducted, the survey objectives must be clearly specified. This can be done by holding consultations with the major stakeholders and considering the intended primary uses of the data. The survey objectives will be the basis for determining the data items to be included in the survey, the domain or the level at which estimates will be generated and the outputs that will be disseminated.

Defining the observational unit and the population unit is essential to the development of the sampling strategy. In agricultural surveys, a farm or agricultural holding is usually the observation unit. According to the FAO World Census of Agriculture (WCA) 2020 definition, which is used by many countries, *"an agricultural holding is an economic unit of agricultural production under single management comprising of all livestock kept and all land used wholly or partly for agricultural purposes without regard to title, legal form, or size"* (FAO, 2015). If the agricultural holding's principal economic production activity is agricultural production, it is considered an establishment within the agriculture industry. Single management may be exercised by an individual, jointly by two or more individuals, by a household clan or tribe, or by a juridical person such as a corporation, religious organization, cooperative or government agency. The holding's land may consist of one or more separate parcels (simple compact blocks of land) located in one or more separate areas or in one or more territorial or administrative divisions, provided that these parcels share the same means of production utilized by the holding, such as labour, farm buildings, machinery or draft animals. In some surveys, the parcel (and not the farm or agricultural holding) is the observation unit. When an area sampling frame is used, in some cases, a tract may be the observation unit.

Survey managers decide how to apply the definition of agricultural holding in the field and in the construction of the sampling frame. In developing countries, the tenurial and cultivation arrangements of agricultural households may change easily. In some instances, a piece of land owned by one farmer can be used as a collateral for a loan in a cropping season, thus giving the lender the right to cultivate the land. A farmer may also plant rice in the wet season and then rent out the land to someone else during the dry season (Colwell, 1989).

The definition of agricultural holding also assumes that the concept of household is clearly understood. In practice, the agricultural holding is usually the same unit as the household as agricultural production operations are interrelated with household management. In applying this definition, it is first necessary to clarify what is meant by land used for agricultural purposes. For example, should it include only land formally classified as agricultural land? Should small vegetable plots in the village be included?

Specifying the outputs at the onset will help in planning and designing the survey, especially in the design of the questionnaire, the data processing system and the survey report. It will also help in ensuring that the data requirements of major stakeholders are met. In developed countries, separate surveys are usually conducted for specific crops for each respective cropping season. In developing countries, because the budget available for agricultural statistics is limited, a single general-purpose agricultural survey is usually conducted on an annual basis. The output required from this general-purpose survey tends to includes total production for major commodities, areas under crops, number of agricultural holdings and their geographic distribution, average size of holding and size ranges, use of fertilizers, irrigation, cultivation practices (multiple cropping, mixed cropping, single cropping), crop calendar, seed variety, and livestock and poultry inventories. Some surveys may also require information regarding the farming household or assessment of the conditions of the land.

In determining the data items that will be included in the survey, the concepts and definitions of these data items should also be specified. Existing approaches for collecting data on major characteristics of interest, such as crop area (planted and harvested) and production or yield, are discussed in the next two chapters. These methods are also explored for various types of cropping patterns (such as mixed crops, root crops and vegetable crops) in the last three chapters of this Handbook.

To plan effective field operations that will minimize non-sampling errors, the following factors need to be considered:
1. the respondents' capacity to provide accurate data;
2. the skill level of field staff or enumerators to implement measurement methods and use technology such as hand-held devices and Global Positioning Systems (GPS);
3. the budget for field operations; and
4. the date on which results should be released.

3.2 DETERMINING A VIABLE SAMPLING STRATEGY

The sampling strategy of a survey consists in the set of techniques applied to select a probability sample and the estimation methods for computing the estimates of characteristics of interest and their corresponding variances. The design of an agricultural production survey with national coverage usually entails a multistage cluster sampling design. Each stage will have a sampling frame that consists of all the sampling units in that stage. For each stage of selection, a sampling frame comprising of all sampling units in that stage of selection must be developed. For example, to implement the first stage of selection, a comprehensive list of all primary sampling units (PSUs) should be constructed from the agricultural census or, if this is not available, from available administrative data. There can be two separate list frames for each developing country – one comprising of all primary sampling units (PSUs) or clusters of small farm holdings and the other, of a list of commercial agricultural enterprises.

The sampling strategy specifies the stages of selection, the sampling unit for each stage, the construction of the corresponding sampling frame, the number of sampling units to be drawn and the rules for assigning the probability of selection. The estimation methods include the determination of the final survey weights that include adjustments for nonresponse and noncoverage and for estimating desired population parameters, such as total crop production, total area and their corresponding measures of precision.

The sampling strategy, including sample size, stages of selection and stratification measures, are determined based on the available budget and the answers to the following questions:
- What are the available sources of data that can be used to construct the sampling frame? Are potential measures of size and information to be used for stratification available?
- What are the crops for which estimates need to be reported? Will there be only one crop or are all crops to be covered?
- At what level of disaggregation will the estimates be reported (for example, at provincial or regional level)? Will production be reported by crop variety or by irrigated or non-irrigated land?
- How is the crop to be estimated distributed across farms? Is it evenly distributed or do only a few farms have it?
- Do the farms producing the crop vary considerably in size? If so, are measures of variability available? Can large units be identified?

3.2.1 Determination of domains

The domain is the level for which separate samples will be planned, designed and selected (Kish, 1987). Examples of domains for surveys having national coverage are provinces, regions or urban/rural domains. The determination of the domain level is an interplay of the survey objectives, the available budget and the desired precision of major estimates. The aim of a general-purpose agricultural survey is to measure major crop and livestock production at a geographic level with a specific political administration, so that policies and interventions can be effectively monitored and implemented at this level.

For example, in the Philippines (see annex 2 to this Handbook, entitled "Country experiences"), the domains for the Palay and Corn Production Quarterly Survey are the rice- and corn-producing provinces. Rice and corn production estimates are reported at the provincial level.

Subdomain level estimates (such as provinces under a region domain or municipalities under a province domain) are not expected to be produced with adequate precision without a large increase in sample size. However, crossclasses that cut across domains (for example, urban/rural with the region as the domain, or irrigated/non-irrigated/upland) and that may comprise one-tenth of the population or more are expected to render reasonable estimates from probability samples.

A factor to be considered in the choice of domains is the possibility of oversampling for certain domains, and therefore the lower precision of estimates that cut across domains. In general, if the population sizes of prospective domains vary, a smaller domain may also require a sample size that will result in a greater selection probability, or what can be perceived as "oversampling". The oversampling of small domains may result in increased sampling errors of other estimates that are not domain-specific. For example, the oversampling of small geographic domains will lead to less precise estimates for subgroups that cut across the geographic domains.

3.2.2 Constructing the sampling frame

A comprehensive sampling frame that includes all population units ensures that with an appropriate selection procedure, each population unit has a chance of being included in the sample and, therefore, robust estimators of population parameters and their corresponding sampling errors can be derived.

In an agricultural sample survey, there are two main types of sampling frames: the list frame and the area frame. Multiple frame surveys that use both list and area frames have also been employed by countries such as Brazil (see annex 2).

The agricultural census, which is conducted every ten years by most countries, is usually the initial basis for constructing the list frame. However, the list frame may incur substantial coverage errors as time passes from the year the agricultural census was conducted. To minimize coverage errors, periodic updating of the list in the sampled areas is essential. Thus, the costs relating to maintenance of the list frame can be high. If the list in the sampled areas are not updated, then the results of the survey may lead to biased estimates.

Area sampling frames are supposed to have complete coverage. However, the initial construction of an area sampling frame can be costly and requires a certain set of skills that may not be present in the statistical offices of developing countries. Sampling areas with minor crops from area sampling frames may also be inefficient. A good example of a survey that uses an area frame is the National Resources Inventory survey of the United States of America.

A multiple frame survey employs both an area frame and a list frame. As discussed above, both frames have strengths and weaknesses. If the list frames and area frames were constructed so as to be able to capture their respective inherent strengths and complement each other so that their weaknesses are reduced, then a multiple frame survey can produce better results.

If simple random sampling without replacement (SRS) is used and there is only one stage of selection, the sample size can be determined based on a tolerable margin of error for estimates of major variables, ε , the variability of the population units and the desired level of significance, α such that $P\left(\left|\bar{y}-\bar{Y}\right| \le \varepsilon\right) = 1-\alpha$. Given these factors and conditions, the sample size will be:

$$n_{SRS} = \frac{n_0}{1+\frac{n_0}{N}}, \text{ where } n_0 = \left(\frac{z_{\alpha/2}S}{\varepsilon}\right)^2 \quad (1)$$

where $z_{\alpha/2}$ is the abscissa of the standard normal distribution at $\alpha/2$, S is the population standard deviation and N is the total number of population units in the sampling frame (Lohr, 2010). If the population standard deviation is not available, it can be estimated on the basis of the sample standard deviation derived from similar existing surveys. If the sampling design is not *SRS*, then the sample size for *SRS* is inflated by the design effect, which will be discussed later.

3.2.4 Stratification and sample allocation

To improve the precision of survey estimates, stratification can be introduced at any stage of selection. To stratify, population units that are similar with respect to a known characteristic of interest (auxiliary data) are grouped together, so that the population variance within groups is small while the variability across groups is large. As the auxiliary data for each population unit in the sampling frame is known, the population proportion of each stratum is also known. That is, the population proportion is $W_h = N_h/N$, where N_h is the number of population units in stratum h that should be defined for all strata.

When stratification is used, the number of samples to be allocated for each stratum can be determined using either proportional or disproportional allocation. Proportional allocation is applied such that the sample size for stratum h is $N_h = nW_h$, where n is the total sample size. Like *SRS*, proportional allocation makes uniform selection probabilities across strata (the Equal Probability Selection Method, or EPSEM). EPSEM ensures that survey weights do not vary widely and therefore that their contribution to the loss of precision is kept to a minimum.

An example of disproportional allocation is the equal allocation model in which $n_h = n/H$, where H is the number of strata, or the square root allocation such that $n_h = n\frac{\sqrt{N_h}}{\sum_h \sqrt{N_h}}$.

When the costs of surveying are known and available and a budget is imposed, the sample can be allocated optimally such that the variance of the sample mean is minimized, subject to a fixed cost. The optimum allocation for stratum h is

$$n_h = n\frac{W_h S_h/\sqrt{c_h}}{\sum_{h=1}^{H} W_h S_h/\sqrt{c_h}} = n\frac{N_h S_h/\sqrt{c_h}}{\sum_{h=1}^{H} N_h S_h/\sqrt{c_h}}, \quad (2)$$

where S_h is the population standard deviation and C_h is the unit cost of surveying in stratum h. Note that if the unit cost of surveying and the population standard deviation are uniform across all strata, the optimum allocation reduces to proportional allocation.

The choice of allocation method depends on the survey objectives. The difference between proportional and equal allocation is substantial when the strata differ markedly in terms of population size. If either one of these allocations is used, it will perform well for its class of estimates; however, it will perform badly for the other class. For example, if only estimates at the domain level will be obtained, then proportional or optimum allocation, depending on the available data, will render the lesser sampling error. However, if subclass comparisons or comparisons between strata will be done (for example, the yield of irrigated versus upland areas or one geographical stratum versus another), equal allocation will provide the more precise estimate.

Because only one allocation can be applied in the survey design and surveys are usually multipurpose, requiring both types of estimates, a compromise allocation that is suboptimal for both classes of estimates but that performs reasonably well for both is often the preferred solution. The Kish allocation (Kish, 1987) offers a good compromise: $n_h = n \frac{\sqrt{H^{-2} + IW_h^2}}{\sum_h \sqrt{H^{-2} + IW_h^2}}$, where I is the index indicating the relative importance of estimates from the proportional allocation. Note that if $I = 0$, Kish allocation becomes equal allocation, while as I becomes large, n_h approaches the proportional allocation. $I = 1$ is used to indicate the equal importance of domain-level estimation and subclass comparison.

3.2.5 Sample selection methods

Probability sample surveys having national coverage usually employ several stages of selection with different sampling units and, therefore, different sampling frames at each stage. The various selection methods that may be used are discussed in detail in annex 3 to this Handbook.

3.2.6 Determination of the survey weights

Following a standard approach, the weights to be used when analysing surveys are developed in three stages. First, base weights are computed to compensate for the unequal selection probabilities in the sample design. Second, the base weights are adjusted to compensate for unit nonresponse. Third, the weights adjusted for nonresponse are adjusted further to ensure that certain weighted sample distributions conform to distributions obtained from another source. These three stages are described further below.

In general, the base weight for a sampled unit is given by the inverse of that unit's probability of selection for the sample. In a domain d, the base weight in an EPSEM sample design will be $w_d = 1/f_d$.

All surveys experience some degree of unit or total nonresponse in which a sampled and eligible unit fails to participate in the survey. For example, the agricultural operator may refuse to participate, or may never be at home at the times the data collector calls. Adjustments are made to the base weights to compensate for nonresponse by sampled units that are eligible for the survey. In essence, the adjustment inflates the base weights of "similar" responding units to compensate for each nonrespondent.

The most common form of nonresponse weighting adjustment is a weighting class adjustment. Weighting classes are constructed on the assumption that the respondents and nonrespondents in the same class are similar. The full sample of respondents and non-respondents is divided into a number of weighting classes or cells and nonresponse adjustment factors are computed for each cell *l* as

$$w_l' = \frac{\Sigma_{i \in rl} w_{di} + \Sigma_{j \in ml} w_{dj}}{\Sigma_{i \in rl} w_{di}} = \frac{\Sigma_{i \in sl} w_{di}}{\Sigma_{i \in rl} w_{di}} \quad (3)$$

The denominator of w'_l is the sum of the weights of respondents (indexed *r*) in cell *l*. The numerator is the sum of the weights for respondents and the sum of the weights for eligible nonrespondents (indexed m for missing) in cell *l*. Together, these two sums in the numerator give the sum of the weights for the total eligible sample (indexed *s*) in cell *l*. The nonresponse weight adjustment w'_l is the inverse of the weighted response rate in cell *l*. Note that the adjustment is applied to eligible units. Ineligible sampled units (such as uncultivated holding, fallow, and units out of scope for a given survey) are excluded.

The weighting cells are created by combining PSUs with similar characteristics in terms of the characteristics of the respondents and the nonrespondents. For example, PSUs in which there is a high proportion of irrigated lands can be lumped into a weighting cell on the assumption that the nonrespondents in these PSUs are similar to the respondents. The combinations are formed within the sampling strata. In forming the cells, consideration is given to achieving a sufficient number of sampled units in each cell and to the response rates in the cells. A low response rate results in a large nonresponse adjustment for the cell, which can significantly lower the precision of survey estimates. Attempts are made to avoid this outcome in the cell formation process.

Generally, weighted sample distributions do not conform to known population distributions (such as the total agricultural area). In particular, sample estimates of population counts generally fall short of true population counts because of noncoverage. Further weighting adjustments – here called population weighting adjustments – may be made to compensate for noncoverage and to make the survey estimates based on the adjusted weights estimates consistent with known population distributions. These weighting adjustments may be made within weighting cells such as the nonresponse cells described above. In this case, the adjustments are often called poststratification adjustments. More broadly, the adjustments may be made using some form of calibration method, such as a raking ratio, generalized regression weighting and linear weighting (Kalton and Flores-Cervantes, 2003).

Ranking is a poststratification adjustment that may be used when the poststrata are formed using more than one variable and only the marginal population totals are known. It makes use of an iterative proportional fitting procedure such that the sample row totals are adjusted to conform to the population row totals; then, adjustments are further introduced such that the sample column totals conform to the population column totals. This process is repeated until the row and column totals both converge to their corresponding population totals.

In generalized regression weighting, weights are adjusted such that weighted sample estimates for quantitative variables conform to the known population parameters. Linear weighting is a special case of generalized regression weighting in which the auxiliary variables are categorical variables. The final survey weight assigned to each responding unit is computed as the product of the base weight, the nonresponse adjustment and the population weighting adjustment as described above. The final weights will be used in all analyses to produce valid estimates of population parameters. The use of the weights in estimation is described below.

3.2.7 Estimating totals and means

Most of the estimates generated from agricultural surveys are in the form of totals, means, proportions or ratios, which are focused on here. However, the methods can be applied more broadly to other types of estimates.

Consider the estimation of a population total. Let y_i and w_i denote the value of variable y and the final weight for respondent *i*. The notation can be applied to a holding or any other unit of analysis. The final weight w_i can be viewed as the number of population units that respondent *i* represents; thus, $\Sigma\, w_i$ estimates the total number of units in the population, *N*. The survey estimate of the population total for variable *y*, denoted by *Y*, is then simply $\hat{Y} = \Sigma\, w_i y_i$.

The estimate $\hat{Y}$ has wide applicability. It can be used to estimate the count of the population with a given characteristic by setting $y_i = 1$ if respondent *i* has the characteristic and $y_i = 0$ if not. For example, with $y_i = 1$ if respondent *i* is farming rice and $y_i = 10$ otherwise, $\hat{Y}$ estimates the total number of holdings planted with rice in the population. The estimator $\hat{Y}$ can be used to estimate the total population with a characteristic in a population subgroup by setting $y_i = 1$ only if respondent *i* is both in the subgroup and has the characteristic. Alternatively, the subgroup total can be estimated by setting $w_i = 0$ for respondents not in the subgroup, or by performing the summation in $\hat{Y}$ only for respondents in the subgroup, that is, $\hat{Y}_s = \Sigma_{i \in g} w_i y_i$, where *g* stands for subgroup *g*.

The extension to estimating a population mean for *y*, $\bar{Y} = Y/N$, is straightforward. With $\hat{Y}$ estimating *Y* and Σw_i estimating *N*, $\bar{Y}$ may be estimated by $\bar{Y} = \Sigma w_i y_i / \Sigma w_i$. Letting $y_i = 1$ if respondent *i* has a given characteristic and $y_i = 0$ if not, $\bar{y}$ estimates the proportion with the characteristic. The estimation of the mean or proportion for a subgroup can be obtained by setting $w_i = 0$ for respondents that are not in the subclass. An example is the proportion of holdings planted to rice, which is the ratio of the holdings planted to rice to the total number of holdings.

3.2.8 Ratio estimation

A further extension is to the estimation of a population ratio of the form $R = Y / X$, where *X* is the population total for another variable denoted by *x*. The ratio *R* may be estimated by means of the ratio estimator $r = \Sigma w_i y_i / \Sigma w_i x_i$. A mean or proportion is the special case with $x_i = 1$ for all units in the population.

To estimate the total area planted to a specific crop, a direct estimate defined by $\hat{Y} = \Sigma w_i y_i$ may be used. However, if there is additional related information available from the survey that is closely related to the characteristic to be estimated, such as the total agricultural area (*x*), and $\hat{X} = \Sigma w_i x_i$, then the total area of the crop of interest can be estimated with $\hat{Y}_R = \frac{\hat{Y}}{\hat{X}} X$, where *X* is the total agricultural area from the most recent census or administrative reporting system. If population total *X*, is not available, then the estimate from the most recent survey can be used. If the population total X is unknown, the sampling error $\hat{Y}_R$ will increase with additional variability associated with the auxiliary variable *X*. The estimate of the variance is usually deemed acceptable if $cv\,(x) \leq 10\%$ and it is tolerable if $cv\,(x) \leq 20\%$. The coefficient of variation (*CV*) of the ratio estimate $\hat{Y}_R$ is lower than that of a direct estimate $\hat{Y}$ if there is a positive correlation between the characteristic of interest *Y* and the auxiliary variable *X*. The *CV* of $\hat{Y}_R$ will be lower than that of the *CV* of $\hat{Y}$ when the correlation coefficient ρ is $\rho > \frac{\mathrm{CV}(x)}{2\mathrm{CV}(y)}$ or when the *CV* of the auxiliary variable *X* is less than twice the product of the correlation coefficient and the *CV* of the characteristic of interest.

With complex sample designs, sample means and proportions of the form $\bar{Y}$ computed for specific subclasses are generally ratio estimators that involve the ratios of two random variables. The denominator of $\bar{Y}$ is Σw_i, which is an estimate of the size of the population or a subgroup, and as such is a random variable. This feature affects the computation of sampling errors for means and proportions, as discussed below.

3.2.9 Variance estimation

The precision of a survey estimate depends not only on the sample size but also on the sample design. There are two main approaches for estimating the sampling errors for survey estimates: a Taylor Series or linearization approach and a replication approach. Both approaches involve approximations and, in particular, both generally assume that the first-stage sampling fractions are small. With this assumption, the sample design can be treated as if the PSUs were sampled with replacement, thus greatly simplifying the computations. Under the "with replacement" assumption, there is no need to explicitly consider the second and subsequent stages of sampling in variance estimation because they are automatically incorporated into the variance estimates.

The following paragraphs provide a brief introduction to variance estimation for estimates from surveys with a replicated sampling design. The Taylor Series Linearization (TSL) approach, which is likely to be used more frequently, as well as the "with replacement" approximation, are presented here. With this approximation, the aspects of the sample design that are relevant to variance estimation are the first-stage strata, the PSUs and the survey weights. Each respondent record in the survey data file must contain information on these three variables to enable computation of variance estimates. The survey sampling error software packages that apply the TSL approach (such as SAS, STATA, R, CENVAR and SUDAAN) require these design variables as inputs.

The identification of these three design variables for the responding sample can be achieved by extending the subscript notation as follows. Let $y_{h\alpha i}$ and $w_{h\alpha i}$ denote the value of variable y and the final weight for respondent i in PSU α in stratum h. As mentioned earlier, this notation can also be applied to holdings or farms as the units of analysis.

Consider the estimation of the variance of an estimated population total. The estimate of the population total is the sum of the estimates of totals for each of the strata: $\hat{Y} = \Sigma \hat{Y}_h$. Under the "with replacement" approximation, the variance of $\hat{Y}_h = \Sigma_\alpha \Sigma_i w_{\alpha i} y_{\alpha i}$ can be estimated by

$$\operatorname{var}(\hat{Y}_h) = \frac{a_h}{a_h - 1} \sum_\alpha \left(y_{h\alpha} - \frac{\hat{Y}_h}{a_h} \right)^2 \qquad (4)$$

where $y_{h\alpha} = \Sigma_i w_{\alpha i} y_{\alpha i}$ is the weighted total of the survey variable y for PSU α in stratum h and α_h is the number of sampled PSUs in that stratum. Note that equation (4) involves computing the totals for each sampled PSU in the stratum and then computing the variance between these totals. All PSUs must be included in the computation of equation (4) even if they do not contribute to the population total (that is, $y_{h\alpha} = 0$).

The calculation of $var(\hat{Y}_h)$ in equation (4) cannot be performed unless at least two PSUs are sampled from the stratum. In designs in which one PSU is sampled from each stratum, the collapsed strata procedure is applied. This procedure involves treating a set of strata as if they were a single stratum with several selected PSUs. The resulting variance estimate overestimates the variance of the survey estimate to some degree. The overestimation is minimized if the collapsing involves pairs of strata with similar means for the survey variables. The collapsing method employed for the full set of PSUs can pair off PSUs within an explicit stratum, putting the first two PSUs sampled from the ordered list into one collapsed stratum and the other two PSUs in the other collapsed stratum.

The extension of equation (4) to variance estimates for estimates of domains and national totals is straightforward. The estimate of a domain total is $\hat{Y}_d = \Sigma_{h \in d} \hat{Y}_h$ and, because of the independence of sampling between strata, its variance is $\Sigma_{h \in d}\ var(\hat{Y}_h)$. The estimate for a national total is $\hat{Y} = \Sigma_h \hat{Y}_h$ and its estimated variance is $\Sigma_h\ var(\hat{Y}_h)$. An estimated variance for a subclass total may be calculated with the same formulae, but with $y_{h\alpha i} = 0$ for units not in the subclass.

As noted earlier, estimates of population means and proportions for the total sample or for subclasses are special cases of ratio estimators. In general, the estimated variance of a ratio estimator r = $\hat{Y} / \hat{X}$ is approximately

$$\text{var}(r) \approx \frac{1}{\hat{X}^2}[\text{var}(\hat{Y}) + r^2\ \text{var}(\hat{X}) - 2r\ \text{cov}(\hat{Y}, \hat{X})] \qquad (5)$$

where $cov(\hat{Y},\hat{X})$ is the covariance of $\hat{Y}$ and $\hat{X}$. The variances $var(\hat{Y})$ and $var(\hat{X})$ can be computed as above and $cov(\hat{Y}, \hat{X}) = \sum_h \text{cov}(\hat{Y}_h, \hat{X}_h)$ may be computed as

$$\text{cov}(\hat{Y}_h, \hat{X}_h) = \frac{a_h}{a_h - 1} \sum\nolimits_\alpha \left(y_{h\alpha} - \frac{\hat{Y}_h}{a_h} \right) \left(x_{h\alpha} - \frac{\hat{X}_h}{a_h} \right). \qquad (6)$$

It must be noted, however, that equation (5) is a reasonable approximation only if $\hat{X}$ has a small CV, for example 10 percent or less.

The estimated variance of a ratio mean or proportion of a subclass can be obtained by setting $x_{h\alpha i} = 1$ for all members of the population belonging to the particular subclass and 0 otherwise, so that $\hat{X} = \Sigma\Sigma\Sigma w_{h\alpha i}$. In the case of a proportion, $y_{h\alpha i} = 1$ for all those with the characteristic of interest belonging to the particular subclass and $y_{h\alpha i} = 0$ otherwise.

The formulae given here can be used to provide estimates for many of the types of estimates that are generated from agricultural surveys. Extensions to other estimates are available in the literature on survey sampling. In practice, however, the variance computations will generally be done using a survey analysis software package. This software will require the stratum, PSU and weight to be identified for all respondents in order to perform the variance estimation calculations.

3.2.10 Forecasts versus estimates

It is important to distinguish between forecasts and estimates. A forecast of yield is determined based on planting intention in some countries and as a prospective yield in advance of crop maturity, while an estimate of yield is made when a crop is mature and ready for harvest. Forecasts of crop production in developed countries are important and methodological researches have been done to reduce forecast error for crop yields. Objective yield measurements during the growing season are conducted to improve the forecast. In developing countries, production estimates of the previous year are used as bases for planning for the current year. Objective yield measurements may be done to improve the production estimates of the current year. Data collection, including objective area and yield measurement, if any, is usually done once a year for some countries and on a quarterly basis for those that can afford more frequent collection efforts.

The FAO publication entitled *Agricultural Market Information System – Crop Yield Forecasting: Methodological and Institutional Aspects* contains a comprehensive discussion of crop yield forecasting.

3.3 SAMPLE DESIGN AND ESTIMATION FOR SURVEYS USING LIST FRAMES

3.3.1 Design overview

A list frame is a list of all sampling units with unique identifiers that can be used to locate those that will be selected. The holdings are usually the ultimate sampling units (USUs). In developing countries, holdings can vary widely in size, some being large commercial holdings while majority are small holdings. Two list frames can be conducted at the onset – one that will cover the small holdings and the other, the commercial agricultural holdings.

The small holdings are usually selected using a multistage sampling design in which each stage of selection will have a different sampling unit; therefore, different sampling frames will be constructed for each selection stage. For example, in a two-stage sampling design where the PSUs are villages or groups of villages and the USUs are holdings (farms), the sampling frame for the first stage will be a list of all villages by domain. The list of farms for each selected PSU will be the sampling frame for the second stage of selection.

A good list frame contains potential measures of size that can be used to select the sampling units. For PSUs, measures of size can be the total area of agricultural land, the total areas of specific crops, the total number of agricultural holdings or the total number of agricultural operators. The PSUs must have unique identifiers and should be easy to locate using the information in the sampling frame. They must have well-defined boundaries. PSUs with overlapping boundaries will introduce bias into the resulting estimates.

The list of commercial holdings or enterprises must include contact details that are current, possible measures of size such as total production or total revenue, and stratification measures such as the commodities produced by each commercial enterprise. There will be additional data collection costs if enterprises in the sample cannot be located.

3.3.2 Construction of PSU sampling frame

The list of commercial holdings or enterprises can be extracted from the agricultural census together with possible measures of size and stratification. The list can then be updated based on information from regulatory agencies under the ministry of agriculture or other government agencies, such as the ministry of trade and industry or the agency responsible for securities and exchange. Other sources of information are organizations of which enterprises are members.

The sampling frame from which small holdings will be selected can also be constructed from the agricultural or population census that also tagged the households in the agricultural sector. The PSU must first be defined. It must have clear boundaries and be easily identifiable, and should have enough elements that could be selected. Examples of possible PSUs are villages or group of villages, enumeration areas defined in the census, or, for some countries, communes. Once the PSU level has been identified, the sampling frame of PSUs is constructed by aggregating the household- or farm-level data to the PSU-level estimates. In so doing, PSUs that are undersized can be linked with adjacent PSUs (collapsed) and those that are oversized can be either subdivided or included in the sample as a certainty stratum. When linking adjacent PSUs, care must be taken to ensure that the newly created PSUs still have clear boundaries and that accurate measures of size and stratification are kept. In the case of oversized PSUs in an EPSEM two-stage sampling design, the number of elements to be selected from the certainty stratum must be adjusted to maintain the uniform selection probability. That is, and $b = f_d x M_\alpha$, where f_d is the selection probability in domain d and M_α is the measure of size of the oversized PSU.

An undersized PSU arises when the number of elements in the PSU is less than b, the number of elements to be sampled. An oversized PSU arises when $P(\alpha) = aM_\alpha \Big/ \sum_\alpha M_\alpha > 1$.

PSUs that are difficult to reach (and thus may require a substantial portion of the limited budget) and those in conflict areas are usually excluded from the sampling frame of PSUs. These usually constitute only a small portion of the population. However, if these PSUs form more than 2 percent of the population, then the weighting adjustments described above may be employed to compensate for the noncoverage. Another issue to be addressed is whether to include PSUs that are classified as urban when they do not encompass any agricultural holdings but may contain the residences of agricultural operators or farmers. If the sampling frame of PSUs was constructed from the population census, which usually lists the locations of the residences of households, it is advisable to include urban areas in the sampling frame of PSUs.

3.3.3 Stratification measures

In a survey for a specific crop, stratification measures that are closely related to the particular crop (such as the percentage of land planted to the crop) can be applied. On the other hand, if the survey is one of general purpose, wherein the production or yield of many crops will be estimated, the choices of stratification measures available from the agricultural census can be the percentage of land that is cultivated or agricultural, as well as geographic divisions, such as districts, a combination of towns, or agro-ecological zones.

Stratification measures should be available for all sampling units in the list frame. For example, in a general-purpose survey, the list frame of PSUs for selecting small holdings should include – for each PSU in a given domain – the measure of size for that PSU, its location (geographic identifiers, including geographic stratification measures, if any), and stratification measures, such as the percentage of cultivated land. Stratification measures would usually come from the same data source from which the PSUs were constructed. However, other data sources, such as satellite imagery or administrative reporting systems, can also be explored provided that these data sources are capable of providing data at the PSU level as constructed. Consistency of PSU definition should be exercised to ensure better, more homogeneous strata. Stratification measures can be used to define strata or, if systematic sampling or systematic probability proportional to size is used, stratification measures can be used to sort the PSUs.

For the list frame of commercial holdings, stratification measures can be the commodity produced by the holding, the size of the holding, or the classification of the commercial holding in terms of revenue, if available.

3.3.4 Determining sample size, sample allocation, PSUs and USUs

After the list frame has been finalized, the sample size can be determined at the domain level using estimates of population variances and the design effects of major variables from previous survey rounds or similar surveys and the desired level of precision. Population variances and intra-class correlations can also be derived from the agricultural census, if available. In a general-purpose survey, there will be more than one major characteristic of interest to be measured and their respective population variances may vary greatly, resulting in a wide range of sample sizes. The most conservative estimate of the sample sizes (largest) may not be cost-effective; therefore, the final sample size must be determined from the range of possible sample sizes on the basis of the budget and infrastructure available. The sample can then be allocated using one of the allocation methods described in section 2.4.

If the average unit costs of surveying a PSU and a holding are available, then the optimum USU size (number of holdings) can be determined using equation (5). For each stratum, the number of clusters can then be computed from the sample size allocated for that stratum and the USU size.

There is no need to determine the number of PSUs and USUs to select commercial holdings because the design does not usually involve cluster sampling.

The measure of PSU size is normally the number of agricultural holdings or the proportion of cultivated land area that is derived from the most recent census. These measures become obsolete as time passes from the census years. The selected PSUs can be relisted to update the measure of size. When this happens, the number of holdings (USUs) to be sampled for the next survey round should be recalculated to maintain a uniform selection probability at the domain level. That is, the updated number of holdings to be sampled, b'_α should be $b_\alpha = b\, N_\alpha / M_\alpha$ where N_α is the updated measure of size and M_α is the original measure of size for PSU α. Unlike the original number of USUs to be sampled, which is uniform across all selected PSUs in a domain, the updated number of USUs can vary across selected PSUs because of the differences in how the PSUs evolve over time. It should be noted that this approach is applicable a general-purpose agriculture survey, in which the shift in agricultural activity is usually captured in other regions that are also represented in the sample. Hence, updating the list of the selected PSUs is sufficient. However, for a particular commodity survey, the shift may be towards the regions that are not well represented in the survey, hence conducting listing operations and adjusting the size of the USUs in sampled PSUs may not be sufficient anymore. In this case, reliable administrative data or other sources like organizations of farmers must be used to update the sampling frame.

Another approach would be to maintain the original number of USUs and recalculate the selection probability such that:

$$P(\alpha\beta) = P(\alpha)P(\beta|\alpha) = \frac{aM_\alpha}{\sum_{\alpha=1}^{A} M_\alpha}\frac{b}{N_\alpha} = f'_{d\alpha} \quad . \tag{7}$$

This implies that the sampling design will not be EPSEM and that selection probabilities and survey weights will vary across selected PSUs.

3.3.5 Estimation

When two list frames are used in a survey, that is, a list frame for commercial farms and another list frame for small holdings, and there is no overlap between the two frames, then weighted estimates of the corresponding totals can simply be added. For example, if the production of a major crop is to be estimated from both frames, such as Y, then

$$\hat{Y} = \hat{Y}^c + \hat{Y}^s \text{ where } \hat{Y}^s = \sum_{d=}^{D}\sum_{h=1}^{H}\sum_{i=1}^{a_h}\sum_{j=1}^{b} w_{dhij}y_{dhij} \text{ and } \hat{Y}^c = \sum_{d=}^{D}\sum_{h=1}^{H}\sum_{k=1}^{n_h} w_{dhk}y_{dhk}, \tag{8}$$

$\hat{Y}^s$, $\hat{Y}^c$ are the weighted total production estimates from the small holding and commercial lists, respectively, w_{dhij} is the final survey weight and y_{dhij} is the crop production for holding j in PSU i, stratum h, domain d, and w_{dhk} is the final survey weight and y_{dhk} is the crop production for commercial holding k in stratum h, domain d. The total crop area *(X)* should be computed similarly; therefore, the yield for a given crop can be derived using ratio estimation such that $\hat{R} = \hat{Y}/\hat{X}$. The variance of the yield estimate will then be given by equation (9).

3.4 SAMPLING DESIGN AND ESTIMATION FOR SURVEYS USING AREA FRAMES

This section presents approaches for designing an agricultural survey that uses an area frame. The Global Strategy's *Handbook on Master Sampling Frame for Agricultural Statistics: Frame Development, Sample Design and Estimation*, provides further details on these methods.

3.4.1 Design overview

An area frame for an area of land such as a state or a country consists of a collection or listing of all parcels of land for the area of interest from which to sample. These land parcels can be defined based on factors such as ownership or, simply, on easily identifiable boundaries. The first elements required to define an area sampling frame are the geographic boundaries of the area of interest. To construct the area frame for a country, its boundaries and a set of rules to divide it into non-overlapping units should be available to ensure completeness and non-redundancy of the frame.

When constructed as described above, an area frame is complete and covers the population of interest. It provides defined boundaries for repetitive surveys and can be used for a considerable length of time without being updated; in addition, it minimizes nonsampling errors with proper photography or maps. However, an area frame becomes inefficient if farms vary considerably in size or some items are rare, such that they appear only in a limited number of farms. It can be said, however, that the same disadvantages apply to a list frame if it does not contain an adequate measure of size.

3.4.2 Construction of the area frame and stratification

The sampling units in an area frame can be points, transects or segments. Points sampling frames are common in forestry but less so in agriculture, although some statistical systems are moving towards this type of area frame. Transects or lines of a certain length are often used for environmental and forestry surveys, to estimate the total length of linear landscape elements. The most common sampling unit is the segment, which is a piece of land defined either by physical boundaries or by a regular geometric shape. The processes of construction of the area frame and stratification vary according to the nature of the sampling units.

In constructing an area frame, mapping materials such as satellite images, orthophotographic images or photo-mosaics are needed so that land areas can be classified by land use or classes of agriculture. Photo-mosaics are composites of contact prints of the same scale that are laid together with a degree of overlap to form a "picture" of a large area, often equivalent to the area covered by a specific topographic map with a scale of 1:50 000 (FAO, 1996; FAO, 2015).

As the crops planted may vary across a domain, the precision of survey estimates could be improved by dividing the land within a domain into strata in which the crops planted are more homogeneous. The stratification of the domain can be done by classifying the land into land-use strata (for example, cultivated land, urban areas and pasture). Each land-use stratum can be further subdivided into more agriculturally similar areas or what could be called segments.

Good stratification ensures precise estimates. Undercoverage or noncoverage errors rarely o in area frames. Example of these errors are when the boundaries exclude certain agricultural areas (such as minor islands) or when some strata are excluded from the sampling. Strata boundaries may be defined using easily distinguished physical features, for example paved highways, roads, railroads, rivers and streams, or by using grids in lieu of physical boundaries.

Within each stratum, segments are constructed so that they are contiguous non-overlapping land areas that divide the survey area. There are three types of segments that are used for an area frame. The most common type are the segments with identifiable physical boundaries: terrain features such as roads, rivers or railroads that are readily available and can provide a clear distinction of the segment. The second type are the square segments defined by grids using map coordinates. The third type are segments defined by a cluster of points.

For segments with physical boundaries, each segment is further divided into non-overlapping tracts. A tract is the land area of a holding that falls inside a segment or the land area of a segment that does not belong to any holding. A tract within a segment is often divided into a number of fields that have distinct boundaries and in which the land use is different.

The reporting or observation unit is the unit to which the information collected about the characteristic of interest is attributed. The observational or reporting unit may be a tract or a holding.

Adjacent segments are grouped into PSUs. Each PSU consists of segments; the number of segments depends on the segment size defined for the stratum and the availability of easily distinguishable boundaries. Segment size may also refer to the number of tracts in the segment or the number of associated holdings or reporting units. While satellite images or photo-mosaics are used to define strata and PSUs, segments are delineated on mosaics of aerial photos that have higher resolution. The boundaries of PSUs and segments within PSUs are digitized with the aid of Global Information System (GIS) software. The total population of segments will be the total of all segments for all PSUs covering the entire frame area.

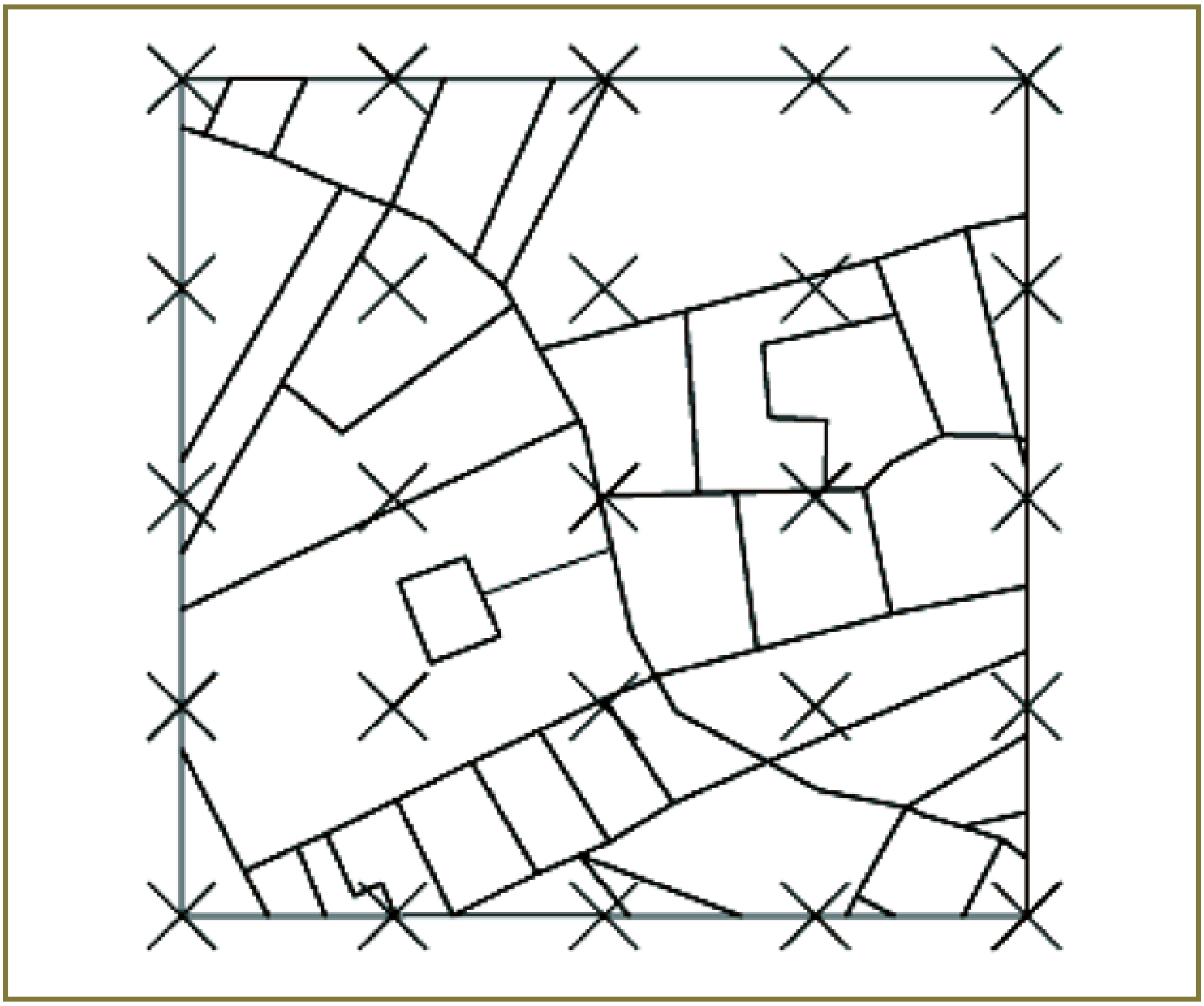
Dividing a stratum into segments can be time-consuming if the boundaries for each segment are defined. To improve the efficiency of this approach, a multistage sampling design can be implemented such that the first-stage sampling units (PSUs) can be larger areas and the second-stage sampling units are the segments of the selected PSUs. Because larger areas are in the first-stage sampling frame, defining the boundaries of the sampling units will not require as much time as the more homogeneous but smaller segments and only the boundaries of the segments of the selected PSUs will be defined.

The second approach for constructing a segment using a regular grid composed of uniformly shaped non-overlapping cells that can be overlaid onto the stratum, as depicted in figure 3.1. Each cell in the grid corresponds to a segment. As the boundaries of the segment are defined outright by the grid, the time needed to construct the sampling frame is greatly reduced. Enumerators may encounter difficulties in the field in locating the boundaries of the selected segments, as they are not visible and mistakes may arise in this process. FAO (2015) noted that such mistakes are reduced when a good GPS instrument is used. Field operations may not also be as straightforward when data that cannot be directly observed (such as use of fertilizers, irrigation practices and seed variety) are required from agricultural holders (Abreau *et al*., 2017), because the agricultural holders in the selected segments must be located and interviewed. Moreover, when there are several holders in the selected segments, the interview time will also be longer.

FAO (2015) cited examples of countries that use point sampling to estimate many types of land use. A cluster of points are sampled in the first stage and points in the selected clusters are drawn in the second stage. The area covered by a cluster or PSU may be formed so as to be approximately the same size as a segment and, therefore, the cluster can be considered an incomplete observation of a segment. This is because instead of mapping and recording all fields in a segment, only a sample of points is recorded. It may be more efficient to cover a larger area for a cluster of points rather than a segment with complete observation including delineated tracts or fields (USU or observational unit).

Gallego (1995) illustrated this approach in figure 3.1. In this figure, 25 points (marked by X) will be observed and recorded to represent a square segment. The enumerator will only document the land use in each of these 25 points instead of digitizing all of the fields in the square grid (indicated by the polygons inside the square). Then, the proportion of area of a particular crop in the segment will be the proportion of points in the square grid that are planted to the crop. Therefore, the areas of the particular crop in the fields in the square segment need not be measured. Results through direct observation are more easily obtained with point sampling, compared to sampling segments.

FIGURE 3.1. POINT SUBSAMPLING FROM A SQUARE GRID.



Source: Gallego, 1995.

3.4.3 PSUs and USUs

Similar to the list frame, the optimum segment size (number of USUs or holdings in the list frame) will be determined on the basis of intraclass correlation and the average unit costs of surveying a PSU and a segment. Following discussion of the general strategy, the estimates done on the basis of a design with more segments in the sample and less holdings will provide more precise estimates than a design with less segments and more holdings. If costs and intraclass correlations are not available, the target segment size can be set as equal to one day of an enumerator's work. Information about the average size of holdings and the percentage of holdings with various characteristics of interests that can be used for stratification are needed for planning purposes and for determining the strata; it may be obtained from the agricultural census or administrative data.

Segments with physical boundaries

PSUs (contiguous groups of segments) are sampled in the first stage of selection; segments are then identified in the second stage of selection. All tracts or holdings are sampled in the selected segments. Their respective boundaries are mapped and digitized.

The ordering of PSUs within each stratum is important to organize the workload of enumerators and to support systematic selection, if implemented. PSUs are given sequential numbers in a serpentine pattern within a small administrative subdivision, such as districts or municipalities, for which agricultural information can be obtained. The ordered list of PSUs within each stratum, together with their assigned measure of size (number of segments), is called the area sampling frame.

These PSUs can still be reordered based on specific agricultural characteristics, thus implementing another stratification level within larger strata. This grouping can be done on paper using the PSU numbers as identifiers; proximity or physical location is not considered a limitation. This reordering or substratification could be denoted in the digitized map using different colour codes.

If segments are identified with reference to permanent physical boundaries, a two-stage sample design is usually applied such that PSUs are selected with probability proportional to size and the segments are selected with equal probability within selected PSUs, thus bringing about an EPSEM selection probability of segments within each stratum.

As previously discussed, replicated samples can be designed from the area frame to add flexibility to the design, as well as facilitate the computation of sampling errors. When the full budget is not available, some replicates may be dropped. On the other hand, without disrupting the general sampling strategy, additional replicates may be added to increase the precision of estimates, if more funds become available. Replicates can also be assigned to different survey procedures, paving the way for testing new methods.

Segments defined using grids

If segments are defined using grids overlaid onto a stratum, they are selected using SRS or systematic sampling within each stratum. While construction of the area frame and sample selection are simpler when grids are used, data collection for a general-purpose survey will not be straightforward because enumerators must also identify all the holders (farmers) who cultivate the selected segments, so that they can be interviewed for the survey. Locating the correct respondents will be as difficult as when segments with physical boundaries are employed, especially in developing countries where the holdings are relatively small and farming arrangements may change frequently from season to season.

Surveys that require field operations may be easier to conduct if points are subsampled inside the selected segments. Box 1 describes examples that subsample points from selected grid cells.

Segments that coincide with the land of the agricultural holdings

To construct segments that coincide with the land of agricultural holdings, a sample of points on a target area is selected. The selected points are identified on the ground and the corresponding holding is selected for the area sample. In developing countries, it may take longer to ascertain the correct holder or farmer for each point as land cultivation arrangements change frequently. It is also for this reason that if the same set of sampled points is to be used for more than one survey round, the holding and farmer must be verified for each survey round. The selection of points can be done systematically or by SRS. This selection technique is often referred to as point sampling.

BOX 3.1. EXAMPLES OF SURVEYS USING AREA FRAMES THAT UTILIZE GRID TO DRAW PSUS AND POINTS TO SELECT SSUS.

The LUCAS (Land Use/Cover Area Frame Statistical Survey) was launched by the European Union (EU) in 2001 based on a nonstratified, systematic two-stage sampling scheme, with PSUs defined as a rectangle of 1 500 m × 600 m following a grid of 18 km (Delincé, 2001). In each PSU, 10 points or SSUs were selected and arranged on two rows of five points each, with a step of 300 m. The point was defined as a circle with a radius of 3 m, for consistency with ground survey specifications. In 2006 (Gallego, 2007), the LUCAS sampling plan became a two-phase sampling scheme of unclustered points (always defined as squares of 3 m).

In 2002, the Italian AGRIT survey (Martino, 2003; Carfagn,a 2007) also adopted a sample design based on unclustered points. It was a single-stage two-phase sampling scheme: the first phase gives a systematic sample of unclustered points that are photo- or imagery-interpreted and the second phase involved subsampled points with higher rates in agricultural strata. A similar approach was tested in Greece in 2004.

The French Ministry of Agriculture conducts the TERUTI, the oldest survey method employing area frames (Gallego, 2015). As a two-stage sample survey of land use and cover, it divides the French territory into a 6 km x 6 km square grid resulting in 15 500 segments as the PSUs. Within each segment, SSUs are defined by 36 points that are 300 m apart from each other and are distributed in a 6 x 6 grid. A total of 550 500 points are sampled each year. A similar framework is followed by the Bulgarian Agro-Statistics Department of the Ministry of Agriculture and Forests in BANCIK (Bettio *et al.*, 2002).

For more a comprehensive discussion of the different types of sampling frames, see FAO (1996), FAO (2015) and Gallego (2015).

3.4.4 Estimation

To derive estimates from an area frame, it is important to verify whether the USU is the same as the reporting unit. If the USU is not the reporting unit, then the base weight, which is the inverse of the selection probability, must be adjusted. For example, if the holding is the reporting unit and the USU is the segment, then the probability of selecting the holding given that the segment has been selected should be incorporated into the selection probability. That is, the probability of selecting holding k in segment j and PSU i will be:

$$P(ijkl) = P(i)P(j|i)P(k|j) \tag{9}$$

The inverse of equation (9) will be the adjusted base weight. Note that even if the design is EPSEM at the segment level, it may not be EPSEM at the reporting unit level. Supposing that the holding was selected with its area as measure of size, the probability of selecting the holding is the area of the holding in the given segment divided by the total area of the segment. The resulting estimator is often referred to as a weighted segment estimator.

If the tract is the reporting unit and the USU is the segment, and all tracts in the selected segment are observed, then the base weight need not be adjusted. This approach is called the closed segment method. Data can be collected through direct observation or through personal interviews with all holders of farms that are in the selected segment. If a holder is not available for the interview, the land area can be estimated from aerial photographs or directly observed in the field. With this method, coverage and non-response errors are kept low. However, when data that cannot be directly observed – such as economic data required for estimating cost of production or data on farm labour, planting intention or crop production – are needed, then the tract may not be a suitable reporting unit and the closed segment method may not render precise estimates.

Another estimation approach is the open segment method, in which the holdings with headquarters inside the selected segment are the reporting units. The headquarters are usually the dwellings of the holders. If this method is applied, then urban PSUs have to be included in the area frame because holders may live in the urban PSUs while cultivating holdings in a rural PSU. This method requires a field operation to be conducted in the selected segment to identify all holders. This process could take longer in urban areas where there are more dwellings to verify. There will also be more cases of non-coverage and non-response with this method, because holders may not be located or may refuse to participate. Adjustments to compensate for non-coverage and non-response may have to be incorporated in the base weight, which is the inverse of the probability of selecting the segment in each PSU.

BOX 3.2. ESTIMATING LAND-USE AREAS USING DOT SAMPLING – A SPECIAL CASE OF A POINT SAMPLING FRAME.

Jinguji (2015) describes the dot sampling method, a special case of point sampling that incorporates the use of Microsoft Excel and Google Earth into the traditional survey method to estimate rice-planted area in Japan. This method has also been implemented in developing countries such as Sri Lanka and Thailand. It enables selection of a sample without requiring a list-based sampling frame to be available. Using Google Earth imagery, sample dots are either randomly or systematically located in the target area. Figure 3.2. shows dots that are systematically marked within the target area. Each sample point (dot) on the ground is identified by its coordinates (latitude, longitude). Software has been developed to generate sample dots on target areas defined on a Google Earth image. The software also lists the sample dot coordinates in a Microsoft Excel sheet, so that they may be easily located during the field survey. Using Google Earth imagery, a desk study of the sample dots is undertaken to identify the dots that fall on cultivated lands and those that do not. Only the dots that fall on cultivated lands are included in the field survey to determine the crops planted. The frequency distribution of sample dots is generated as per land use categories of interest, including crop area. The estimate of land use under each category is determined by the product of the corresponding percentage share (based on frequency distribution) and the total survey area. This simple approach could be applied to estimate main crops as well as minor crop areas and types of non-agricultural land.

FIGURE 3.2. SAMPLE DOTS SYSTEMATICALLY LOCATED ON A TARGET AREA DEFINED ON GOOGLE EARTH IMAGERY.

Source: Jinguji, 2015.

Jinguji found that land use estimates obtained through the dot sampling method were closer to official estimates. However, as Google Earth employs images captured on different dates with varying resolutions, the delineation of various crops may have some inconsistent results. Guidelines should be established to address cases in which the images are updated and not geometrically consistent.

After the final survey weight has been determined, estimation methods for totals, means and ratios as discussed in sections 2.7 and 2.8 above can be employed.

3.4.5 Application of area frame surveys in developing countries

Although the area frame ensures minimal non-coverage error and more efficient stratification and estimates, it requires the use of technology such as GIS and access to satellite images and orthophotographs that may be beyond the capacity of the NSO staff. While, in recent years, the cost of satellite imagery has remarkably declined and training programmes on the use of GIS and remote sensing have become readily available, the use of these new technologies in developing countries is still in its infancy. The discussions of these topics provided in this Handbook only give an overview of how an area frame can be constructed and utilized. However, many issues that require more advanced and technical concepts, have not been fully discussed here.

It is important to note that the development phase of the area frame needs skilled staff that can devote most of their time to frame construction and field verification, as well as a substantial budget, which could be beyond what the government in question can afford. While the area frame can be constructed with technical assistance from the development community, its sustainability and usefulness can only be guaranteed with strong and continuing government support.

3.5 SAMPLE DESIGN AND ESTIMATION FOR SURVEYS USING MULTIPLE FRAMES

3.5.1 Design overview

Multiple frame sampling involves the joint use of two or more sample frames. For the purposes of agricultural surveys, this usually involves an area frame and a list frame. Both of these frames have inherent strengths and weaknesses. In multiple frame sampling, both types of frames are strategically employed, such that their strengths are optimized and their respective weaknesses are overcome. To illustrate, a list frame with measures of size can be efficient for sampling purposes. However, a list frame usually suffers from non-coverage error. An area frame may be complete but inefficient where measures of size are needed for sampling purposes. Therefore, a list frame could be constructed for large or rare holdings, while an area frame could be used to select general items that are capable of minimizing coverage errors.

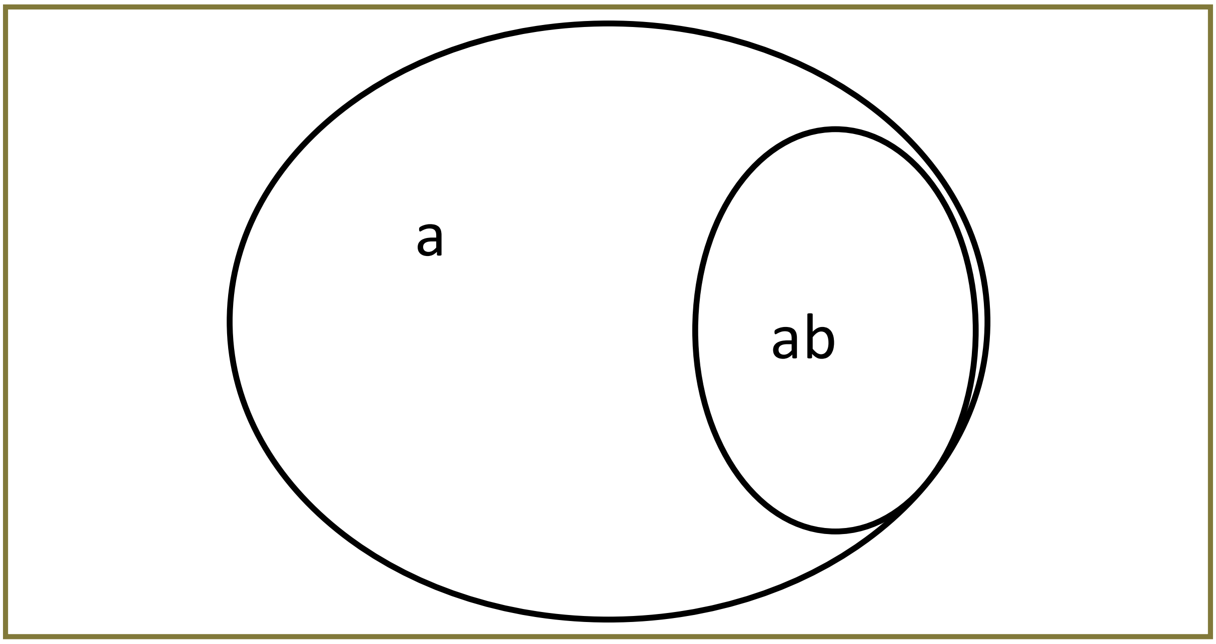
3.5.2 Construction of frames

Area and list frames can be constructed as described in the previous sections. The list frame can be constructed from various sources, such as agricultural censuses, organizations of farmers or commercial enterprises, and information from agricultural extension workers and regulatory agencies. The list frame should contain information that could assist the location and identification of large or rare holdings, as well as those that could be used as a stratification measure, such as total land area and crop areas. This list should be updated before major surveys are conducted. The definition of "large holding" should be determined at the onset of list frame construction. An analysis of the distribution of holdings using the latest agricultural census could form the basis of this decision.

For the multiple frame design approach to be effective, two basic assumptions must be satisfied: (1) the combined sample frames must represent the target population; and (2) it must be possible to easily identify the provenance of each population unit (or the frame(s) to which it belongs).

The two surveys – one that uses the area frame and another that uses the list frame – may be conducted independently; the results may be combined at the analysis phase to derive the overall estimates. However, the association between the sampling unit and the reporting unit must be examined for the two sampling frames. The sampling unit for the area frame is a segment that is a unit of land. Depending on the estimation method used, the reporting units could be tracts (smaller areas within a segment) or holdings. For the survey using a list frame, the sampling unit and reporting units are holdings. The sampling units in the list frame are expected to be covered in the area frame. Figure 3.3 illustrates this association as a Venn diagram in which the big circle (a) represents the area frame, and the small circle represents the list frame.

FIGURE 3.3. ILLUSTRATION OF MULTIPLE FRAMES FOR AGRICULTURAL SURVEYS.



Source: Author elaboration, 2018.

3.5.3 Estimation

There are two types of estimators that could be used to derive the overall estimates – the screening estimator and another estimator that incorporates the degree of the overlap into the overall estimates. If the screening estimator is used, overlaps should be determined before initiating the estimation procedures for the survey. The overlap between these two frames must be determined by matching the names of the holders, which will only be available for the survey using the area frame after the survey has been conducted. If the name of the holder (operator) of a tract of land in a selected area frame segment is also on the list frame, an assumption is made that the same land would be reported if the name had been selected from the list frame; therefore, for this operation, the two frames overlap. Duplicate observations should be removed. The duplicate holdings are counted only under the survey using the list frame.

The equation to calculate the estimator for an overall crop production will then be similar to equation (12). That is,

$$\hat{Y} = \hat{Y}^A + \hat{Y}^L \text{ where } \hat{Y}^A = \sum_{d=}^{D} \sum_{h=1}^{H} \sum_{i=1}^{a_h} \sum_{j=1}^{b} w_{dhij} y_{dhij} \text{ and } \hat{Y}^L = \sum_{d=}^{D} \sum_{h=1}^{H} \sum_{k=1}^{n_h} w_{dhk} y_{dhk} \tag{10}$$

$\hat{Y}^A$, $\hat{Y}^L$ are the weighted total production estimates from the area frame and list frame, respectively; w_{dhij} is the final survey weight; y_{dhij} is the crop production for segment j in PSU i, stratum h, domain d; w_{dhk} is the final survey weight; y_{dhk} and is the crop production for holding k in stratum h, domain d. It should be noted that when the weighted segment method is used to estimate Y^A, the final survey weights and thus the formula will change to:

$$\hat{Y}^A = \sum_{d=}^{D} \sum_{h=1}^{H} \sum_{i=1}^{a_h} \sum_{j=1}^{b} \sum_{m=1}^{c_b} w_{dhijm} y_{dhijm} \qquad (11)$$

where w_{dhijm} is the final survey weight and y_{dhijm} is the crop production for holding m, segment j in PSU i, stratum h domain d. The final survey weight should include the adjustment to the base weight resulting from the difference in the reporting unit (holding) and the USU (segment).

The second type of estimator does not require that overlap be identified and excluded from the area frame. Instead, the sampling weights assigned to the overlapping units will correct for the multiple chances of being sampled. This estimator addresses the problem that arises when overlaps are not easily distinguished. The names of holders in the list frame may change and, without the georeferenced location of the dwelling units of holders in both frames, it may be difficult and costly to ascertain the overlap.

Given that the estimator of the characteristic of interest using an area frame only is $\hat{Y}^A$, using a list frame only is $\hat{Y}^L$, and the estimator for the overlap is $\hat{Y}^{A,L}$, there are several estimators that fall under the second type of estimator, three of which are the following:

- Hartley's general class of dual frame estimators $\hat{Y}_H = \hat{Y}^A + p\hat{Y}^{A,L} + (1-p)\hat{Y}^L$, where p is the value that will render a minimum $Var(\hat{Y}_H)$.
- The Fuller-Burmeister estimator $\hat{Y}_H = \hat{Y}^A + \hat{Y}^L + p_1\left(\hat{Y}^{A,L} - \hat{Y}^L\right) + p_2\left(\hat{N}^{A,L} - \hat{N}^L\right)\hat{Y}^L$, where p_1 and p_2 are estimated so that they minimize the $Var(\hat{Y}_{FB})$.
- The Skinner-Rao estimator, which is a simple weighted combination of the characteristics of interest, a pseudo-maximum likelihood:

  $$\hat{Y}_{SR} = \frac{N^A - \hat{N}^{A,L}_{SR}}{\hat{N}^A}\hat{Y}^A + \frac{\hat{N}^{A,L}_{SR}}{p\hat{N}^{A,L} + (1-p)\hat{N}^L}\left[p\hat{Y}^{A,L} + (1-p)\hat{Y}^L\right]$$

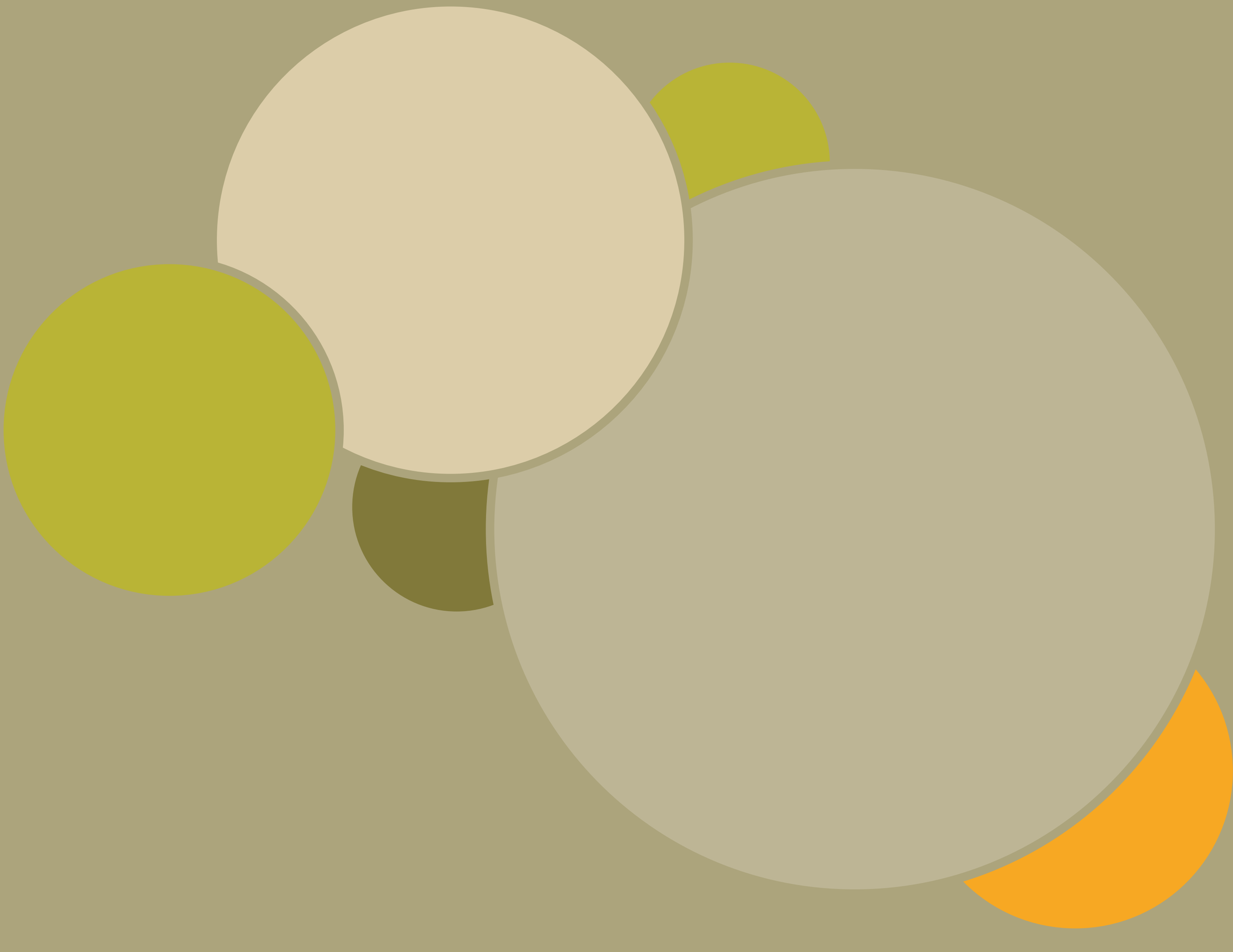
  where $\hat{N}^{A,L}$ is the sum of the weights of the overlapping sampled units and $\hat{N}^L$ is the sum of the weights of the sampled units in the list frame; $\hat{N}^{A,L}_{SR} \cong p\hat{N}^{A,L} + (1-p)\hat{N}^L$ p is determined so that it minimizes the $Var(\hat{Y}_{SR})$.

For more information regarding the estimation of multiple frame surveys, please refer to the Handbook on Master Sampling Frames for Agriculture (GSARS, 2015).

The screening estimator is simpler in form compared to the second type of estimator that has an overlap component. The variance of the estimator is usually lower. However, it requires identification of overlap at the onset, before the analysis phase. This screening process is costlier. On the other hand, the estimator with overlap does not require screening, although it does need a more technical and lengthier analysis.

The total crop area $(\hat{X})$ should be computed in a similar manner. Thus, the yield of a crop can be derived using a ratio estimation such that $\hat{R} = \hat{Y}/\hat{X}$. Assuming that all duplicate observations have been eliminated, the variances for the total crop production will be the sum of the variances of the estimates from the area frame and list frame surveys, which are estimated using equation (8). Similarly, for the variance of the total crop area, the covariance between the total production and total area estimates can be computed using equation (10), while the variance of the yield estimate can be computed using equation (9).

Although multiple frame sampling can sharply reduce the variance of many estimates, it should not be considered in all cases. If procedures are not followed with care, non-sampling errors will greatly exceed gains in precision. Also, multiple frame sampling is not efficient when the list of large operators (holders) or those with rare crops is not comprehensive.



4

Methods for measuring crop area

Crop area, together with crop yield and production, is a key variable collected in agricultural production surveys. Crop area and production are inputs used in deriving the crop yield, which can be used as a measure of productivity. The total area that is planted with various crops is also needed by the government for planning and monitoring purposes.

In some developing countries, the area and production for each crop are collected, while for others, the crop area and yield derived from objective area measurements are recorded. Countries that use list frames would usually collect the area and production of crops from the agricultural holders. Those that use area frames are more to likely collect the area and yield of crops. In the case of the former, the yield of crops can be computed at the domain level, while the production of crops can be derived from the available area and yield data. In both of these approaches, the crop areas must be measured; this underscores the need for accurate crop area estimates.

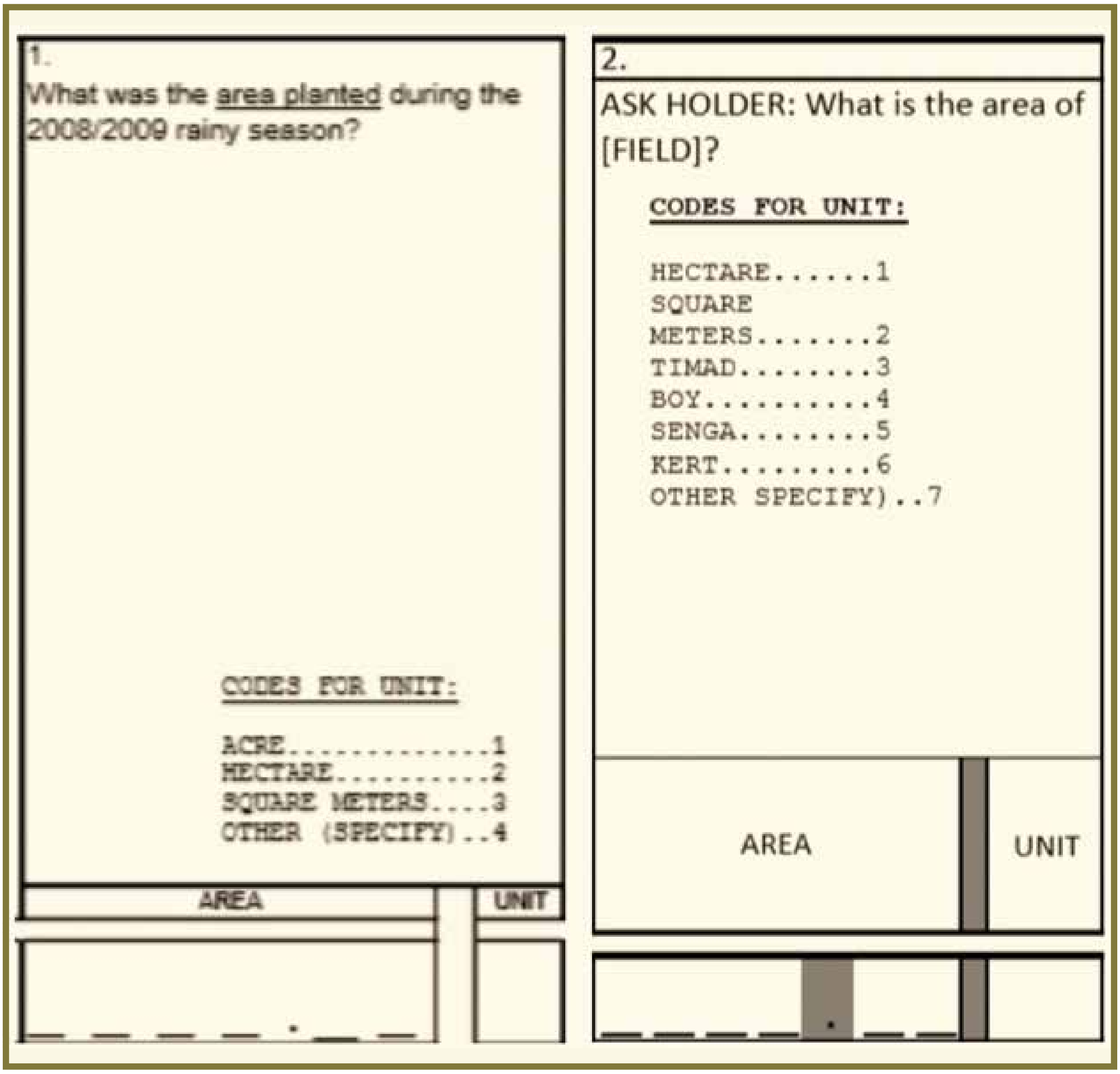
Several methods may be used to measure crop area: he farmer declarations, objective area measurements (including GPS area measurements), other land surveying methods, and remote sensing. This chapter discusses these methods, their implementation, important considerations, their respective advantages and key issues.

Throughout the chapter, reference will be made generically to the concept of crop area, which, as extensively reported in the Glossary (annex 1), corresponds to the area shown on cadastral maps that is grown with crops (FAO, 1982). However, certain variables are of particular interest when producing crop statistics. More specifically, when collecting information at the crop level, as a minimum requirement, data on the area planted and the area harvested should be produced. In addition, measuring or asking about the area irrigated, the area applied with fertilizers and the area applied with plant protection products (PPPs) by crop could offer additional information on the agricultural practices adopted in a given country.

4.1 METHODS BASED ON FARMER DECLARATIONS

Reference to farmer declarations is the most commonly used method for collecting data on crop area. Because it does not require measurement tools nor longer field operations, it is inexpensive. If this method is used, the budget allocated for data collection can cover a larger sample. In this method, a farmer or holder is asked, by means of a personal interview, for the total area he or she cultivates. Because smallholders in developing countries tend to plant a variety of crops in several small parcels, he or she is requested to identify the area of each parcel. The enumerator and the farmer may visit all parcels that the holder cultivates to estimate the surface area by visual inspection (David, 1978) or the farmer may simply enumerate them from memory.

FIGURE 4.1. QUESTIONS FOR FARMER ASSESSMENT OF PLANTED AREA.


1.
What was the area planted during the 2008/2009 rainy season?

CODES FOR UNIT:
ACRE..............1
HECTARE...........2
SQUARE METERS....3
OTHER (SPECIFY)..4

AREA | UNIT

2.
ASK HOLDER: What is the area of [FIELD]?

CODES FOR UNIT:
HECTARE......1
SQUARE METERS.......2
TIMAD........3
BOY..........4
SENGA........5
KERT.........6
OTHER SPECIFY)..7

AREA | UNIT

Sources: Malawi Third Integrated Household Survey, 2010/11 (left); 2011/12 Ethiopian Rural Socioeconomic Survey (right).

In obtaining farmer estimates of the crop area, the question is stated clearly and directly. The questionnaire would usually require the area and unit of measurement of each crop for each parcel of the selected holding. For example, the questionnaire developed by GSARS for AGRIS asks, for each crop cultivated in the agricultural holding, to report the total area planted and harvested in specific units of measurement that are customized to the country context.

Although the use of farmer declarations is the simplest of all methods, several factors are considered to maximize the accuracy and usability of the collected data. Land area can be measured using standard units, non-standard units or both. The suitable choice of units largely depends on the country setting. Whenever non-standard units are used, suitable conversion factors must be available. The conversion factors should be specific to the location of the field for cases in which non-standard units have similar names but different measures. Conversion factors are important, as they make the data usable by transforming the reported areas to a single unit of measurement. These conversion factors are usually developed by the country's ministry of agriculture or NSO or by the proponent of the survey, prior to the conduct of the survey. Familiarity with measurement units could also improve the accuracy of farmers' reported areas, as is the case in Nigeria and Ethiopia. Indeed, it was found that farmer estimates were more accurate when reported in non-standard units (Goundan and Domgho, 2016; Carletto *et al*., 2016b). An example of a non-standard unit is the timad, a unit of measure used in Ethiopia that is equal to the land area that an oxen pair can cultivate in one day. Depending on the texture and moisture of the soil, slope of the plot, and the oxen's strength, a timad may correspond to different sizes (Carletto *et al*., 2016b).

The accuracy of farmer declarations and memory recall has been investigated in several validation studies and country experiences by comparing them with objective measurements. In the Philippines, two studies report that farmers overestimated the relevant area by 6 to 8 percent. While there is a high correlation between the measured areas and farmer-reported areas, a third study showed that farmers generally underestimated the crop area (David, 1978).

These opposing results could be explained by the consistent outcomes obtained in other countries, in which it was showed that small parcels are generally overestimated by farmers while large parcels are underestimated (Ajayi and Waibel, 2000; Goundan and Domgho, 2016; De Groote and Traore 2005; Carletto, Savastano and Zezza, 2011; Carletto, Gourlay and Winters, 2013). A similar trend may be observed at the farm level, where smallholders tend to overreport areas and large landholders, instead, underreport them (Carletto *et al*., 2013). In southern Mali, farmers were found to overreport the area of plots smaller than one ha, while the rate of underreporting increases as the plot size increases. A regression analysis of the observational error showed that every ha increase in plot size is underreported by one-third ha (De Groote and Traore, 2005). Rounding also increases the bias in farmers' reported areas. It was found that farmers tend to report areas rounded off to the nearest acre or half-acre (Carletto *et al*., 2011), or nearest quarter- or third-ha (David, 1978).

Some farmers may not trust the enumerators to keep their responses confidential and may fear that the information they provide would subject them to higher taxation. Moreover, farmers with limited formal and agricultural education may not be able to give accurate information (De Groote and Traore, 2005). Carletto *et al*. (2016b) revealed that the accuracy of farmer estimates improves when household heads are literate but degrades in relation to the years of education, possibly as a result of the lesser time spent on in agricultural activities, thus leading to difficulties in approximating the crop area. An analysis of LSMS-ISA data gathered in Nigeria and the United Republic of Tanzania showed that older household heads are more biased when reporting plot areas. Specifically, in Niger, Burkina Faso and Uganda, older respondents have the tendency to overreport crop areas (Carletto *et al*., 2011; Goundan and Domgho, 2016).

The accuracy of farmer estimates could be increased if the area of the field had been previously measured. The farmer estimates of cotton fields were found to be more accurate than those of cereals in southern Mali. This is possibly because previous measurements had been done by a cotton company, and because the application of fertilizers and pesticides in cotton fields require area measurements (De Groote and Traore, 2005).

When farmers or agricultural holders are unaware of the area of their holdings, they approximate the area based on seed rate or the wages of planters or harvesters. In selected provinces of the Philippines, the correlation between seed rate and objectively measured area ranges from 0.60 to 0.93, making seed rate a possible concomitant variable that could be used to improve the accuracy of farmer estimates of crop area. Its accuracy could also be improved if both reported areas and objective measurements could be drawn from at least one subsample of the sample, where a correction factor may be constructed based on their correlation (David, 1978). The bias in farmer estimates could be reduced by asking enumerators to accompany farmers in the visual inspection of plots. Despite entailing greater costs, this improves the accuracy of estimates, as it is ensures that all fields are accounted for and that enumerators are capable of identifying and discussing discrepancies in farmers' self-reported estimates (De Groote and Traore, 2005).

4.2 METHODS USING LAND SURVEYING TOOLS

4.2.1 Traversing (compass-and-rope) or polygon method

This method is known by several names: the traverse measurement, traversing, chain-and-compass, the Topofil method or the polygon method. It is one of the most common traditional methods used to measure crop area (MAC, 1965; Schøning *et al*., 2005). It requires basic geometry skills, a compass, a measuring tape, ranging poles, two to three persons, and a computational tool. Compared to farmer assessment, this method renders more accurate measurements of the crop area.

To measure a specific crop area, the boundary of the plot is first cleared of all obstructions. The farmer and enumerator walk around the boundary of the plot to delineate the area to be measured. In the process, the enumerator places markers: ranging poles, that visibly identify the corners of the plot. The markers or poles serve as the vertices of the polygon that covers the crop area. All of the line segments defined by two consecutive poles will then be measured, as well as the angles inscribed by two consecutive line segments. These will be the inputs into the calculation of the area described in detail in annex 4 to this Handbook.

The traditional procedure of evaluating the area of a field on the basis of measurements entails plotting the field in the office with a ruler and a protractor and then measuring the area of the sketch by using a planimeter or grid paper. The more recent approach employs a handheld device, such as a tablet, in which the measurements of distance and bearings are keyed into an application that directly computes the crop area and closing error (Carletto *et al*., 2016a). If the plot has an irregular shape or is bounded by curved edges, the crop area will be bounded by several line segments or sides. The number of sides of the polygon increases as the plot becomes more irregular in shape. Defining more than ten sides of the polygon, however, leads only to a minimal improvement in the measurements. In practice, to approximate curved boundaries, protruding pieces that are left out are balanced by including other small pieces that are not part of the plot. This give-and-take process could result in errors and may cause a gap in the side of the polygon. This is known as the closing error, or the distance between the starting and last corners of the polygon that the enumerator failed to connect. The measurement should be repeated when the closing error exceeds 3 percent of the plot's perimeter (Casley and Kumar, 1988).

Selecting the plot corners is especially difficult for plots with irregular shapes. An increase in the demarcated plot corners may increase bias, as minimal errors in measuring a single corner may lead to a large measurement error when they are pooled across all corners of the plot. Also, repeating the measurements when there is a large closing error increases the time required for measurement (Carletto, 2016b).

For this method to provide accurate crop areas, the enumerators must be well-trained on the polygon method and must exercise due diligence to ensure that the closing error is kept at a tolerable level. The time required for measurement and computations are longer. Because of these factors, the viable sample size for the agricultural survey may provide good area estimates only at the national level, and not at the subnational level.

The various methods to measure distances are discussed in annex 5. All of them may be used to measure the length of the line segments of the polygon, as well as the sides of the rectangles and triangles used in the method discussed in section 4.2.2.

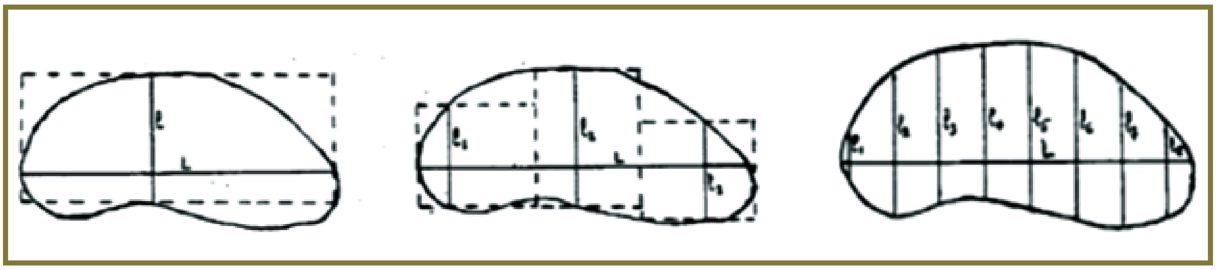
4.2.2 Rectangulation and triangulation

The rectangulation method divides the area to be measured into imaginary rectangles. The enumerator walks across the crop area to measure the length and width of the imaginary rectangles using a Trumeter wheel or a measuring tape. The areas of the rectangles are then computed and pooled together to give the total plot area (Fermont and Benson, 2011). Figure 4.2 illustrates the rectangulation method.

Similar to the rectangulation method, triangulation divides the entire crop plot into triangles. The enumerator crosses the whole field to mark the vertices of the triangles and measures the base and height of each triangle using a Trumeter wheel or a measuring tape. The areas of all triangles are computed and pooled together to provide an estimate of the total plot area (Fermont and Benson, 2011). Figure 4.3 illustrates the triangulation method.

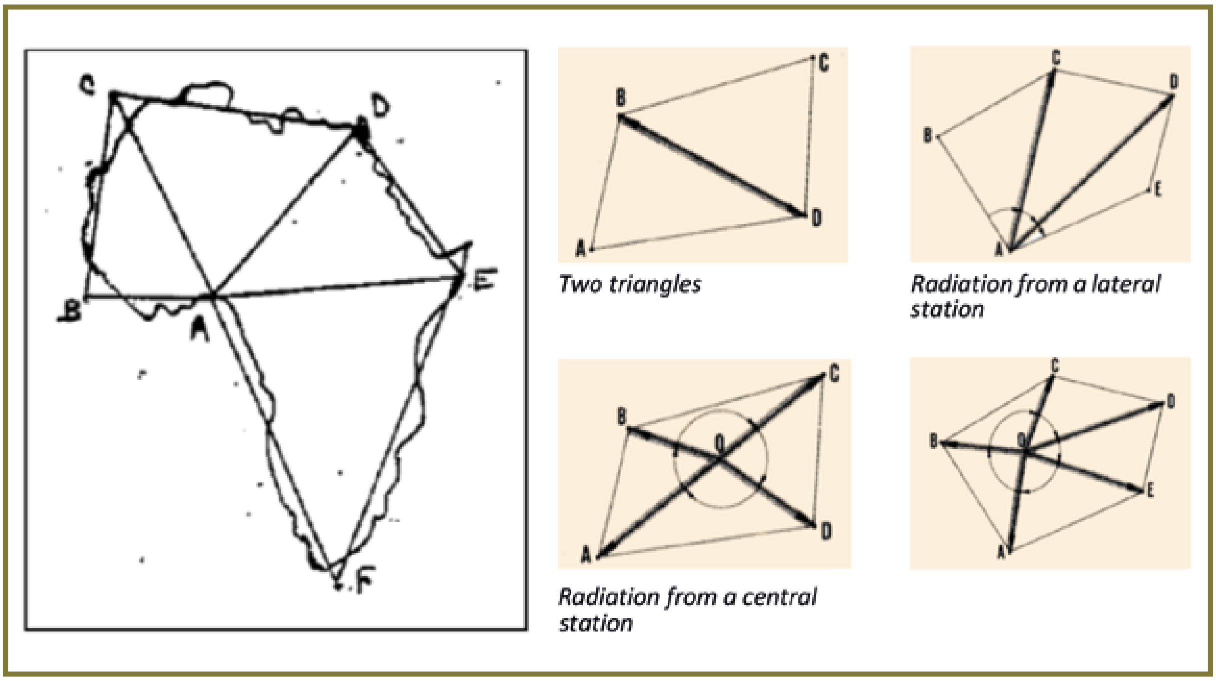
The rectangulation and triangulation methods require less time, enumerator skills and resources to implement, compared to the polygon or traversing method described above. However, neither of these methods will be straightforward to implement if the crop is tall and lush, making it difficult for the enumerator to visualize the division of the crop area. Farmers may not also allow the enumerator to walk across the crop area to avoid damaging the crops.

FIGURE 4.2. SUBDIVISION OF LAND INTO RECTANGLES.



Source: FAO, 1982.

FIGURE 4.3. SUBDIVISION OF LAND INTO TRIANGLES.



Source: FAO (1982; 2006).

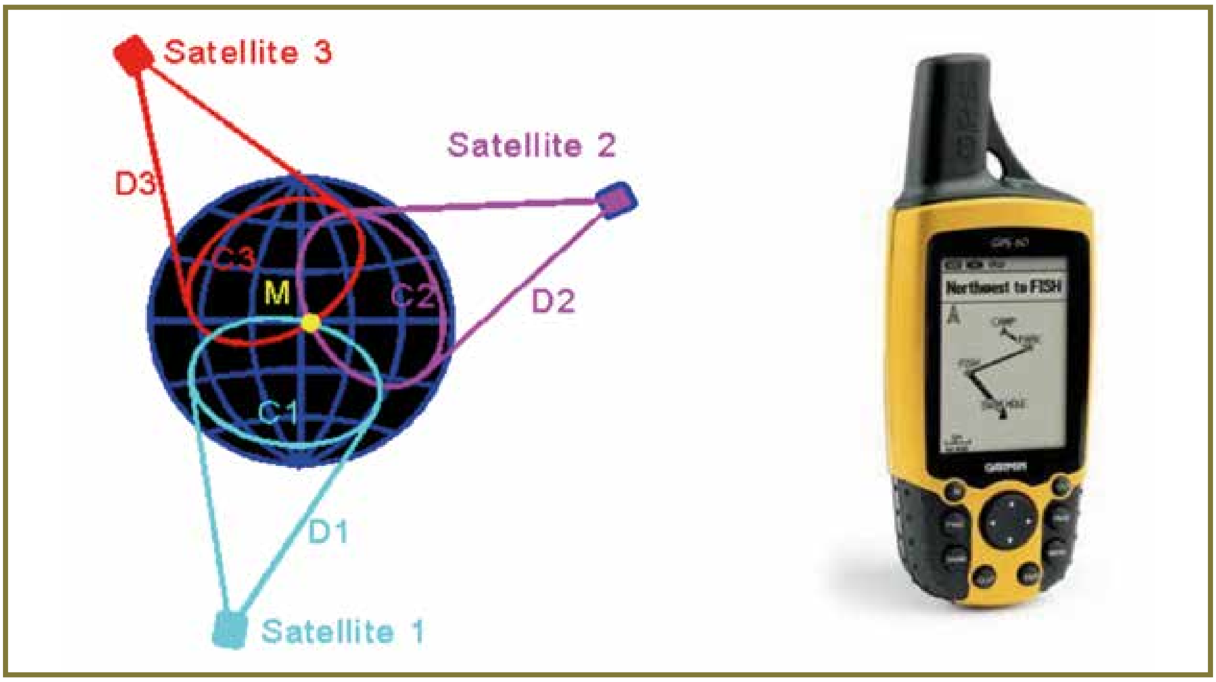
A quick way to check the calculation results of the triangulation and rectangulation methods is to us the P^2/A method. If the perimeter is known, the area of a field can be estimated easily, by dividing the perimeter by a number between four and five and then squaring the result (FAO, 1982). The choice of divisor is subjective and depends on the degree of complexity of the boundary, such as the number of its sides or concavities. Thus, if the field shape is complex, the divisor should be closer to five, whereas if the field shape is approximately square or rectangular, then the divisor should be closer to four. When the field is highly variable, the divisor might even exceed five.

The P^2/A method, where P stands for the perimeter and A for the area, is a subjective method based on a relatively stable relationship between the squared perimeter of a field and its area (Mpyisi, 2002; Fermont and Benson, 2011). This method provides an indicative figure that could be used to compare the results of the rectangulation or triangulation methods. A large discrepancy should alert the enumerator to check the calculations. While the P^2/A method enables estimating a field area quickly, it should not be considered a viable method for measuring crop area that can be used in isolation, as there is no proven mathematical relationship between the perimeters and areas of complex polygons.

4.2.3 Global Positioning System (GPS)

A GPS is a navigation system that employs orbiting satellites to provide location and time information anywhere on Earth. A GPS device delivers the geographic coordinates (*x- and y-axes*) and elevation (*z-axis*) of a location. Although it was initially used to determine the location of a point, technological advancements have made it possible for GPS devices to also determine elevation and land area. These developments have led the GPS to become a significant tool in crop area measurement, with the additional benefits of requiring less time and labour (Sud *et al*., 2015).

FIGURE 4.4. USE OF GPS DEVICES TO MEASURE CROP AREA.



Source: Author elaboration, 2018.

A GPS device could automate the polygon method described above. Similar to the polygon method, the enumerator and farmer must initially pace the perimeter of the plot to be measured to identify and clear the boundary. Starting from a chosen corner of the plot, the enumerator crosses the boundary again, this time activating the measurement feature of the GPS device. The enumerator briefly pauses at plot corners to enable the device to capture the point locations. The measurement is completed when the enumerator has paced the entire boundary of the plot and has returned to the starting point (Carletto *et al*., 2016a). Lines are defined by connecting successive points stored in the device. A polygon is obtained when the starting and final points connect. If the enumerator fails to connect these points, they are automatically connected by the device. The area is computed by GPS based on the resulting shape. Incorrect area measurements could arise from an "impossible surface", resulting for example from an overlap in traversing the boundary. For example, crossing over while pacing the plot boundaries resulting in an "8" shape leads to a computed area equal to the difference of the area of the two circles (Keita, Carfagna and Mu'Ammar, 2009).

Keita *et al*. (2009) provides a series of steps to be followed when using GPS, including before and after the measurement process:

- the GPS device should be tested before use;
- the acquisition strategy should be designed in advance;
- the material required to check data in the field, such as maps containing coordinates and validation points, should be prepared in advance;
- the available data or maps should be uploaded into the receiver;
- fields to be measure are identified;
- the borders of the identified fields should be marked;
- the data are stored and reported, either by recording them on paper or in the GPS device; if GPS is used for storage, the coordinates of the field borders are also stored;
- the shape of the plot shown on the screen of the device during the test and during the acquisition of operational data should be observed to avoid errors;
- a post-acquisition integrity check is advisable to avoid redundant control points; and
- post-acquisition control – which includes checking the authenticity of the data, computing the perimeter/area ratio and visualizing the plot borders – should be conducted and inserted into the virtual globe.

The following issues should be considered when using GPS to measure crop area.

Selecting the GPS device. Devices with GPS and GLONASS compatibility are recommended as they may access more satellites, leading to improved accuracy and acquisition time. For operations in the United States of America, Europe or Asia, position is determined with improved accuracy by devices compatible with the local augmentation system. Some common features of GPS devices are direct area measurement, plot outlines storage, cameras, and altimeters (Carletto *et al*., 2016a). The FAO studies conducted to provide a scientific basis for recommending the use of GPS for crop area measurement revealed that some types of GPS perform better than others under different conditions. The selection of a GPS device should be done with great care; operational use, battery life, robustness under difficult field conditions and simplicity of use should be taken into account (Carfagna and Keita, 2009).

Testing the device. Manufacturers often provide a measure of the positioning errors of GPS devices; however, they do not indicate the precision of area measurements. Prior to data collection, the receiver, its settings, and measurement protocols should be tested in the field to determine the device's precision. Validation of the device also provides information on ease of operation, battery life when used in the field, merits and drawbacks of the device model, and its compatibility for use in data collection, considering aspects such as metadata and its data downloading and storing features (Keita *et al*., 2009).

Battery storage. Securing a sufficient power supply is one of the major problems faced when using a GPS device for measurement. Some GPS devices may be plugged into an automobile's 12-volt power port. Certain GPS units use a large-capacity battery, while others use an external box of dry-cell batteries. Small GPS units typically save weight and space by using two AAA cells; however, they tend to last only four or five hours on one set of batteries. Larger handheld devices use up to four AA cells (which means a greater weight), although these may run for as long as 12 hours per set (Sud *et al*., 2015).

Positioning errors. The GPS receiver computes a satellite's distance based on the difference in the time between the satellite's signal transmission and the receiver's signal reception. The measured distances from at least one satellite are used by the GPS receiver to calculate the position of the operator. The longitude, latitude, elevation and time at a given location are determined by the GPS receiver by accessing four satellite signals. The accuracy of GPS receivers is subject to atmospheric conditions, the position of satellites above the Earth's surface, and signal barriers such as mountains and buildings. Further, GPS receivers are more accurate when access to more satellites and with angled distances of satellites that are far from each other are available (Keita *et al*., 2009). During the time required to measure the area of a plot logging along the perimeter, the GPS constellation can be considered relatively stable, such that positioning errors do not affect the measurement. Area measurement error is linked to operator speed and to the acquisition rate of the GPS device. Field area measurement errors may be limited by selecting an appropriate combination of operator speed and GPS acquisition rate; for parcels up to 4 ha in size, the "optimum" range of speeds for operators on foot is between 0.5 m/s and 2 m/s (1.8-7.2 km/h) (Bogaert, Delincé and Kay, 2005).

There may be erroneous readings in GPS measurements resulting from the interference of trees or projection problems in hilly areas (Schøning *et al*., 2005; Sempungu, 2010). It should be noted, however, that the latter is a problem for any method measuring crop area on steep slopes. This is related to the fact that the measured crop area should not be the physical area measured on the ground, but rather its projection onto a horizontal plane (Muwanga-Zake, 1985). The projection problem may become acute on slopes greater than 10 degrees. As the introduced error is 0.4 percent and 1.5 percent for slopes of 5 and 10 degrees, respectively, the projection problem may be ignored on slopes of less than 10 degrees (Fermont and Benson, 2011). Therefore, it is recommended to use clinometers with a compass and rope in hilly regions with steep slopes.

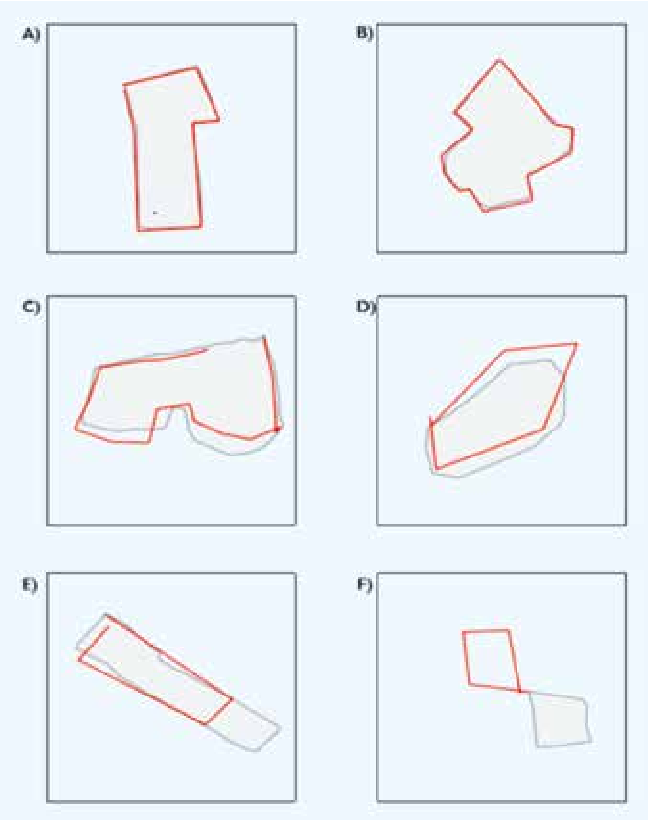
Questionnaire design and additional questions. Carletto *et al*. (2016a) recommends that questionnaires be designed such that information can be easily and consistently transferred from the GPS device to the questionnaire. The questionnaire must reflect the settings and display of the GPS device, specifying the desired number of decimal places for measurements in degrees and minutes. Items requiring GPS measurements should be consistently indicated throughout the questionnaire. Furthermore, additional items such as the number of acquired satellites, plot slope, the surrounding canopy cover, and cloud cover should be recorded as indicators of the GPS measures' precision.

Selecting plots. Plots are commonly located separately from the household. At the planning stage, protocols on selecting which plots to measure using GPS should be drawn to manage the time and costs of the survey. These guidelines are based on the distance of the plots from the household. Such guidelines should be formulated with care. Excessive restrictions on which plots to exclude from GPS measurement may burden enumerators with long travels to measure a single plot. On the other hand, guidelines that easily exclude plots for measurement could lead to a great deal of missing data. In Malawi, enumerators are instructed to consult their supervisors when plots are two hours away (on foot) from the household. Furthermore, parcels that are far from the households but are located close together are grouped and measured. In the United Republic of Tanzania, enumerators are required to measure plots that reachable by foot or any form of transportation no more than one hour away from the household. In Uganda, instead, only parcels located within the EA are measured (Carletto *et al*., 2016a).

As the results of empirical research continue to affirm the high level of accuracy of GPS measurements, GPS has already been used in some studies as a benchmark for investigating the performance of farmer assessments (Carletto *et al*., 2013; Goundan and Domgho, 2016), whereas the compass-and-rope method has traditionally been used as the benchmark for examining other methods (Carletto *et al*., 2016a).

The use of GPS proves to be a practical method in measuring crop area as the prices of GPS devices become more affordable and their precision and reliability improve. However, a crucial issue affecting GPS is the selection of plots to be measured, which gives rise to missing measurements whenever the plots are not immediately accessible (Carletto *et al*., 2011). In large-scale surveys employing GPS, 15 percent to 30 percent of plot measurements are often missing. One way of controlling the bias and gaps in GPS measurements is to impose proper field guidelines. It is also advisable to collect farmer estimates of crop area together with GPS measurements (Carletto *et al*., 2016b), as these have been found to be good predictors of GPS measurements and can be used in the imputation of missing values (Killic, Zezza, Carletto and Savastano, 2013).

FIGURE 4.5. PLOT OUTLINES OBTAINED USING THE COMPASS-AND-ROPE METHOD (RED) AND GPS (BLACK).



- A and B show that there is a small bias between the two methods
- C shows a large plot with 19 sides and with a non-negligible closing error
- D shows a simplified outline using the compass-and-rope method, where the number of vertices is only half of the number of vertices in the GPS outline
- E shows an incorrect identification and simplification of the target area using the compass-and-rope method
- F results from incorrect recording of bearings, that is, front and back bearings have been switched

Source: FAO (1982; 2006).

4.3 METHODS USING REMOTE SENSING

Remote sensing technology has been used in agricultural statistics. In general, remote sensing is the measurement or acquisition of information of a given property of an object or phenomenon through a recording device that is not physically or closely in contact with the object or phenomenon under study. The recording devices could be aerial cameras, lasers and radio frequency receivers, radar systems, sonars, seismographs, gravimetres, magnetometres and scintillation counters (FAO, 1982). Remote sensing technology has greatly improved over the years, from sporadic aerial photography to frequent high-resolution satellite imagery, from black-and-white film coverage to multispectral digital scanners (Craig and Atkinson, 2013).

Remote sensing has three broad uses for agricultural statistics. It could be used for stratification in developing an area sampling frame. Graphical images can guide field operations. Remote sensing could also be used to estimate crop area at a higher level of disaggregation. The use of remote sensing in developing an area frame is discussed in chapter 3 of this Handbook. More detailed discussions are available in the *Handbook on Master Sampling Frame for Agricultural Statistics: Frame Development, Sample Design and Estimation* (GSARS, 2015). This section provides an overview of the use of remote sensing to estimate crop area.

Crop area can be measured from remote sensing images, which are representations of parts of the Earth's surface as seen from space. The images may be analogue or digital. Aerial photographs are examples of analogue images, while satellite images acquired using electronic sensors are examples of digital images. A digital image comprises of a two-dimensional array of individual picture elements, called pixels, arranged in columns and rows. Each pixel represents an area of the Earth's surface. A pixel has an intensity value and a location address in the two-dimensional image.

To estimate the area of a specific crop at the level of an administrative area, such as a district or province, the administrative boundary of the target area is overlaid onto the digital image. The land use of each pixel within the administrative area is then identified. The pixels that have the specific crop are counted. The proportion of these pixels is obtained and multiplied by the total administrative area to derive the direct estimate of the area of the specific crop. This direct estimator is often biased and is capable of providing reasonably good estimates only if the pixels are clearly identifiable (Carfagna, 1999), such as when the area of irrigated crops in a very dry region is evaluated from a digital image taken during the summer.

To derive an accurate crop area estimate using the approach described above, the pixels must be correctly distinguished. Pixels may be classified using a supervised or an unsupervised approach, or a combination thereof. In a supervised approach, the pixel classifier system has the advantage of domain knowledge that can be used to learn the relationship between the data and the classes. The number of classes and the prototype pixels for each class can be identified using this prior knowledge. When there is no access to domain knowledge or analyst experience is lacking, the data can still be analysed such that pixels are grouped into clusters based on similar statistical properties. When prior knowledge is available for some classes and not for others, or when the prior knowledge is only known for some dates in a multidate set of satellite imageries, then the supervised and unsupervised approaches can be combined to classify the pixels. The different techniques that can be applied to implement these approaches are discussed in detail by Defourny in the *Handbook on Remote Sensing for Agricultural Statistics* (GSARS, 2017; chapter 2).

The domain knowledge used for supervised classification may also be referred to as training samples, and could be data collected through fieldwork or from high-resolution aerial photographs and satellite images. Training samples are selected on the basis of the spatial resolution of the remote sensing data used, the availability of ground reference data, and the complexity of the landscapes in the study area.

When the training samples are collected from the field, they are usually referred to as ground truth data. The ground truth data set includes geographical location (village, district or province), crop name, coverage, condition, whether irrigated or rain-fed, expected yield, and sowing and harvesting dates, among others, along with a close-view and wide-view photograph of the field (Ray and Neetu, 2017). They are collected from selected locations within the study area using probability sampling techniques.

When the ground truth data set is compared with the image classification results, the pixels may be correctly identified or they may be misclassified. A confusion matrix, an example of which is given in table 4.1, may be constructed. In the example, there are k classes or categories that are used to classify the pixels. These could be type of land use (crop land, commercial, water, residential, forest, etc.) or types of crops. The cell value n_{ij} represents the number of pixels that were identified as class i using satellite imagery but were found to belong to class j using ground truth data (reference data). The total n_{i+} is the number of pixels classified as class I based on the satellite imagery, and n_{+i} is the pixel count for class i based on reference data. The diagonal values represent the pixel counts that were correctly classified, for each class.

TABLE 4.1. CONFUSION MATRIX TEMPLATE.

Classified data	Reference data					Total
	Class 1	Class 2	Class 3	. . .	Class k	
Class 1	n_{11}	n_{12}	n_{13}	. . .	n_{1k}	n_{1+}
Class 2	n_{21}	n_{22}	n_{23}	. . .	n_{2k}	n_{2+}
Class 3	n_{31}	n_{32}	n_{33}	. . .	n_{3k}	n_{3+}
Class *k*	n_{k1}	n_{k2}	n_{k3}	. . .	n_{kk}	n_{k+}
Total	n_{+1}	n_{+2}	n_{+3}	. . .	n_{+k}	n_{++}

Source: Author elaboration, 2018.

Three measures of accuracy are usually derived from the confusion matrix. These are producer accuracy, that indicates how well a specific class was classified; user accuracy, that shows the degree of likelihood that a specific class truly represents that category on the ground; overall accuracy, which is the proportion of correctly classified pixels and the Kappa coefficient, which measures the agreement between the classification map and the ground truth data. Usually, when producer accuracy and user accuracy rates are reported, the rates of omission and commission error are also presented. Omission error occurs when a pixel for a specific class is incorrectly assigned to another class. Commission error happens when a pixel under a specific class is identified as belonging to a different group.

As an example, the direct estimator of the area of class i is n_{i+}/n_{++} *x total study area*. The commission error for class i will be $\sum_{j=i+1}^{k} n_{ij} \Big/ n_{i+} \times 100$, and the omission error for class i will be $\sum_{j=i+1}^{k} n_{ji} \Big/ n_{+i} \times 100$.

The corresponding user accuracy for class I is $n_{ii}/n_{i+} \times 100$, and the corresponding producer error will be $n_{ii}/n_{+i} \times 100$. Overall accuracy will be $\sum_{i=1}^{k} n_{ii} \Big/ n_{++} \times 100$.

The Kappa coefficient of agreement is: $\hat{K} = \dfrac{n_{++}\sum_{i=1}^{k} n_{ii} - \sum_{i=1}^{k}(n_{i+} \times n_{+i})}{n_{++}^2 - \sum_{i=1}^{k}(n_{i+} \times n_{+i})}$.

In some cases, when the overall accuracy is high and some individual classes may present substantial error. If users are interested in certain classes, it is important that producer and user accuracy, as well as the omission and commission errors of these classes, are examined.

When there are budget or time constraints on ground surveys or when target areas are not accessible, high-resolution imagery with image interpretation may be substituted for ground-truth data. While past studies have shown that the results obtained with this method are less than satisfactory (Atkinson and Craig, 2013), Marshall *et al.* (2011), on the other hand, demonstrates that in Niger, the visual interpretation of high-resolution satellite imagery of a sample of grid points can be used instead of ground-truth data. This result implies that it is possible to use a visual interpretation of points sampled using probability sampling techniques on high-resolution imagery over large areas as ground truth, so that the relationships observed from the "ground truth" substitute and the medium-resolution satellite imagery can be extrapolated over an entire country.

While this technique may be more cost-effective than one that requires a ground survey, the omission errors for crop detection, and especially for irrigated crops, may be substantial (Atkinson and Craig, 2013). Research has shown that commission and omission errors in visual interpretation do not balance out and could often result in large bias of the crop area estimates.

When a ground survey is conducted, the results of the survey and those of the analysis of satellite images can be combined to improve the accuracy of area estimates. There are two types of estimators that can be employed in this process: calibration estimators and regression estimators. Calibration estimators use remotely sensed data as an auxiliary variable in the estimation process. The commission and omission errors of a confusion matrix are then used to correct for the estimate's bias. In the direct calibration estimator, user accuracy (and commission error), is used while the inverse calibration estimator employs producer accuracy (and omission error). For a more comprehensive discussion of these estimators, readers may refer to the annex of the *Handbook on Remote Sensing for Agricultural Statistics* (GSARS, 2017).

A regression estimator may be used to derive an estimate of a crop area for a large area of study if there are ground observations and corresponding calculations from satellite images. As an example, suppose that the large area of study is divided into L strata and that there is a set of sampled observations (point sampling or segment) in each stratum, for example the area of crop c (Y) and the corresponding area of crop c as measured on satellite imagery (X). Then, a linear regression model $Y = a + bX + \varepsilon$ can be applied such that for a stratum h, the regression estimator $\hat{y}_{h,reg} = \bar{y}_h + b_h(\bar{X}_h - \bar{x}_h)$ can estimate the area of crop c, where $\bar{y}_h$ is the average ground-reported crop area per sample segment of stratum h, b_h is the regression coefficient of the ground-reported area on the remote-sensing-derived area based on n_h segments for stratum h, $\bar{x}_h$ is the average remote-sensing-based crop area per sample segment of stratum h and $\bar{X}_h$ is the average remote-sensing-based area for all frame units of stratum h. The total area of crop c for the entire study area will be $\hat{Y}_R = \sum_{h=1}^{L} N_h y_{h,reg}$, where N_h is the number of population units in stratum h. Population units can be segments, grids or points. Survey weights should be properly incorporated when calculating the estimates from ground-truth data.

This approach has been extensively studied in the United States of America: in its June Agricultural Survey, the crop data are collected on a sample basis through the questionnaire method. Other examples of the application of the regression approach may be found in potato and canola-rapeseed crops in Canada and wheat crops in Brazil, where using Landsat MSS (Moriera, Chen and Batista, 1986); in both countries, a Landsat Multispectral Scanner (MSS) is used. Gonzalez-Alonso and Cuevas (1993) achieved better performance when the regression estimator was used with confusion matrix information over a test site in Spain. Gonzalez-Alonso, Soria and Cuevas-Gozalo (1994) compared results of the regression approach using two different sampling approaches: square sample segments and irregular segments over a test site of 900 km2 in Spain. The weighted relative efficiency was higher when the square-segment approach was used than when the irregular segment approach was adopted.

While the United States and many other countries use regression to improve crop area estimates achieved with remote sensing, procedures based on confusion matrices and calibration as discussed above are also possible. The choice between calibration and regression estimators depends on the characteristics of the ground-truth data. Gallego (2011) recommends calibration estimators if the ground-truth data were sampled as unclustered points and regression estimators when data are measured from segments or clustered points.

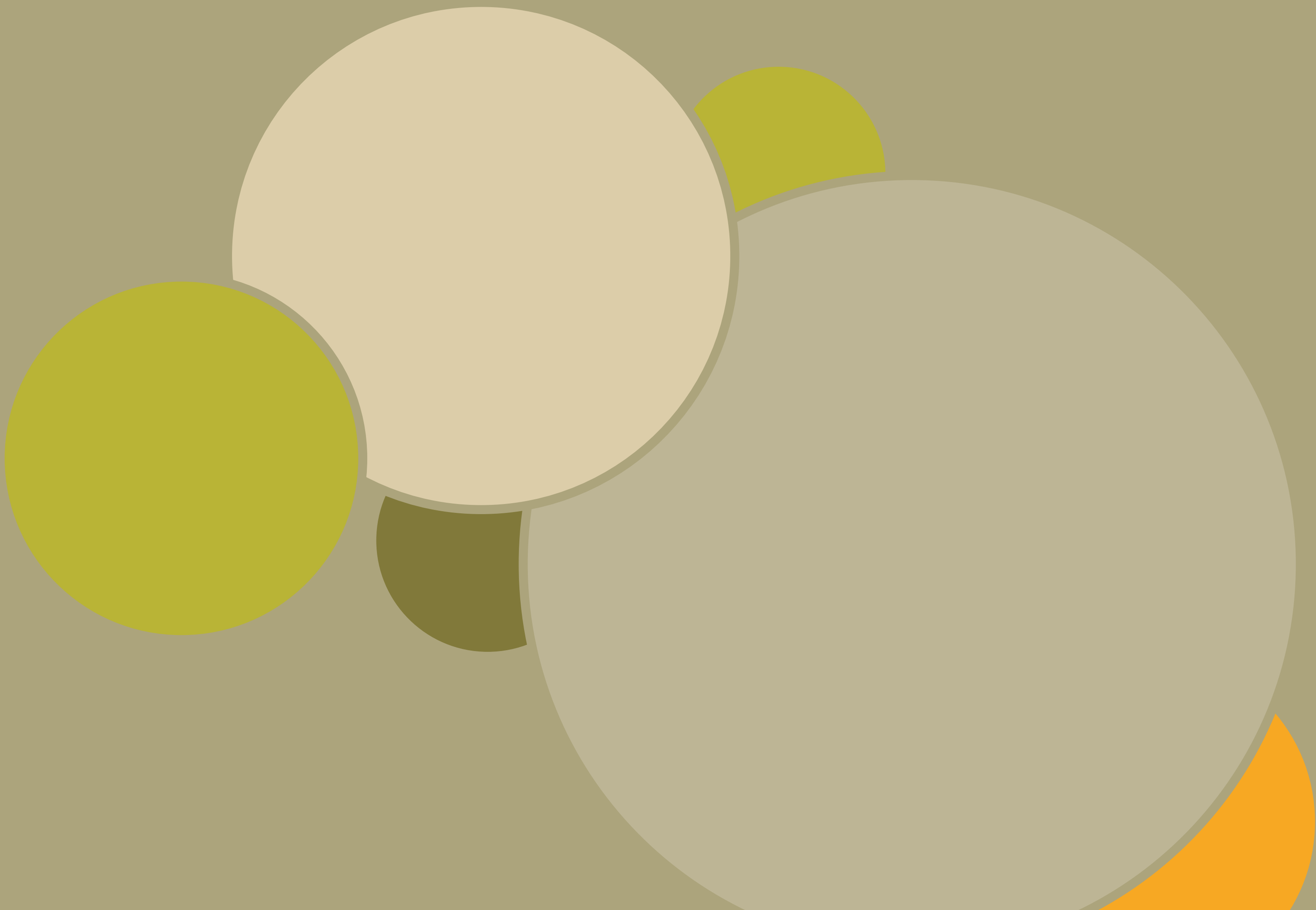
New technologies are constantly being developed. Therefore, methodological research must maintain a brisk pace if it is to improve crop area estimation, which is important for monitoring food security. It is also equally important that procedures to adapt these methods to developing countries, which face limitations in terms of infrastructure and staff resources, be studied.

4.4 COMPARISON OF EXISTING METHODS FOR MEASURING CROP AREA

To determine the appropriate area measurement technique to implement, it is necessary to balance several factors: the target degree of accuracy, the available budget for the agricultural survey, the skill level of the enumerators and supervisors, the size and shapes of the parcels, and the type of crops covered in the survey. Table 4.2. summarizes the various measurement techniques discussed in this chapter. It may also be possible to combine these techniques. For example, if the available budget cannot cover the GPS measurement for all samples, then all of the sample's farmers or agricultural holders may be asked to provide their respective crops' areas, and the GPS measurement could be done only on a randomly selected subsample. If designed properly, this approach is capable of providing measurement error estimates that can be used to adjust, if warranted, the crop area estimates to improve their accuracy.

TABLE 4.2. COMPARISON OF EXISTING METHODS FOR MEASURING CROP AREA.

Method	Cost-effectiveness	Scale	Estimate precision, biases, errors	Ways to improve the estimation
Farmer estimates	Cheap and fast	Plots to landscape	Fairly accurate. However, at times farmers may deliberately provide overestimations, underestimations or not have the information required.	• Apply correction factors with a known covariate, such as seed rate. • The enumerator should visit the plots with the farmer.
Traversing or polygon method	Cheap; however, time-consuming and cumbersome to implement	Plots to farm	Highly accurate when implemented with the necessary diligence. Traditionally used as a benchmark for crop area measurement.	This method should be implemented by skilled enumerators equipped with computational tools.
GPS	Costly; however, saves time	Farm to landscape	Highly accurate for large plots, with decreasing precision as plots get smaller. Underestimates plots with complex shapes.	• Select the appropriate combination of operator speed and GPS acquisition rate. • Secure an ample power supply for the GPS. • Impose proper field guidelines to control bias and missing plot measurements.
Rectangulation and triangulation	Tedious and time-consuming	Plot level to field	Reliable estimate for regularly shaped plots	• Increase the number of rectangles / triangles but will trade off time efficiency.
P^2/A (Perimeter squared over Area)	Quick and straightforward, if the perimeter is known	Field level	Subjective; estimates are rarely acceptable because no mathematical relationship is established between the perimeter and the area	• The enumerator must take note of the ratio to be used by inspecting the shape of the plot. • Must not be used as a standalone method.
Remote sensing	Expensive, although the cost of satellite imagery is decreasing. Quick and straightforward area estimation.	Higher level of disaggregation	Fairly accurate, especially for vast geographical areas, if higher resolution images are used. Accuracy decreases for smaller plots.	• Ground-truth data must be collected to evaluate accuracy and, if warranted, improve the crop area estimates.



5

Methods for measuring crop production or yield

Crop production and yield are important productivity indicators that are used as bases for planning and monitoring at various levels of government and by several sectors of society. Statisticians also use these indicators as inputs when compiling national accounts statistics, such as gross value added for the agricultural sector. This chapter reviews several of the methods available for measuring crop production or crop yield. The choice of data collection method depends on the available resources, the scale of investigation and the level of precision desired. The resource requirements, as well as the advantages and disadvantages of each method are presented in this chapter.

The nature of the crop must also be considered when choosing the data collection method. Certain crops, such as cereals, have well-defined cropping seasons and are harvested within a limited period. Other crops, such as bananas, roots and tubers have staggered maturity and may be harvested throughout the year, making the implementation of some data collection methods difficult. Moreover, smallholders tend to practice mixed cropping or intercropping. While measuring production in this case may be straightforward, defining crop area and, consequently, crop yield may be complicated. The next three chapters of this Handbook provide more detailed procedures for all of these situations. Chapter 6 discusses methods for measuring crop production and yield for mixed crops, Chapter 7 deals with methods for measuring root crops and chapter 8 focuses on vegetable crops.

Definition of crop yield

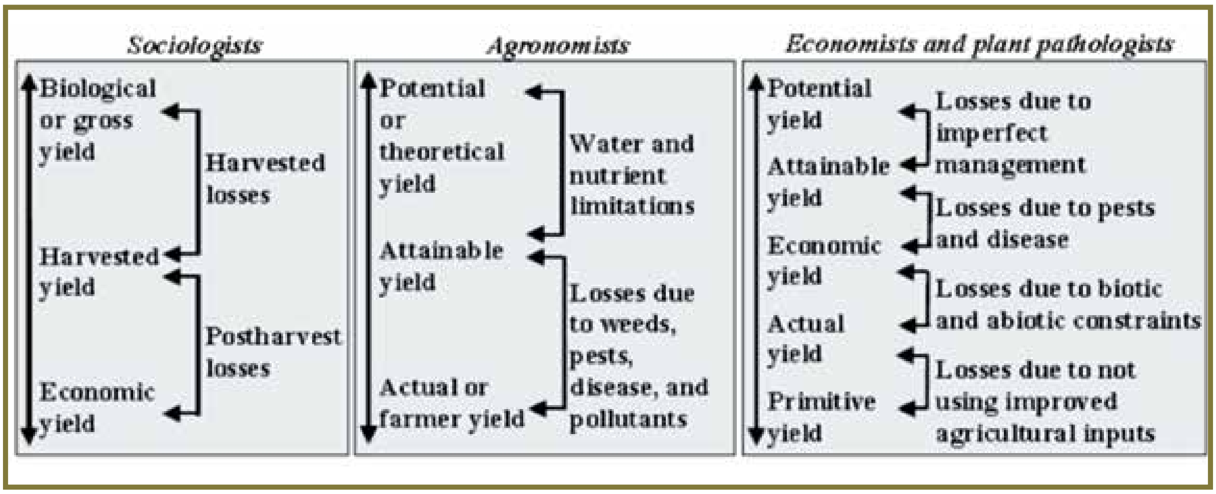
In any crop yield estimation, there are two primary components: crop production (total quantity of farm produce) and area harvested, which must be initially estimated. Crop yield is then defined as crop production divided by area harvested. The unit of measurement is often kg or metric tonnes (MT) per ha.

$$Yield\ of\ crop\ X = \left(\frac{Quantity\ harvested\ in\ mt}{Area\ harvested}\right)$$

If there are many farms on which the area planted and area harvested differ significantly from one another, the calculated yield could be multiplied by the area planted to calculate the potential production to estimate the loss of production, which is given by the potential production minus the actual production (GSARS. 2018).

Several disciplines, including sociology, agronomy and economics, as well as plant pathology, have proposed various definitions, concepts and approaches, and even different terminologies, to discuss crop yields. Economists and plant pathologists appear to share similar ideas. Figure 5.1. summarizes the general theoretical crop yield concepts used by sociologists, agronomists, economists and plant pathologists (Fermont and Benson, 2011).

FIGURE 5.1. SUMMARY OF THEORETICAL CROP YIELD CONCEPTS USED BY SOCIOLOGISTS, AGRONOMISTS, ECONOMISTS AND PLANT PATHOLOGISTS.



Source: Fermont and Benson, 2011.

Among the four disciplines, many countries commonly use the sociologists' concepts of crop yield, which are described below:

- Biological yield or gross yield is the yield obtained before consideration of any losses during and after harvest.
- Harvested yield is the yield obtained after considering harvesting losses, for example, losses due to shattering during harvest of grain (Poate, 1988).
- Economic yield is the yield obtained after considering losses due to post-harvest operations such as cleaning, threshing, winnowing and drying (Casley and Kumar, 1988; Keita, 2003).

The following sections discuss the various methods for estimating crop production or yield.

5.1 COLLECTION OF PRODUCTION DATA THROUGH PERSONAL INTERVIEWS

The production data collection mode of many agricultural surveys and censuses conducted in developing countries is the personal interview, because of the limited budget available for data collection. Depending on the time of interview, the sample farmer or holder is asked to estimate their production or to recall the amount that they have produced for each parcel, field or farm they cultivate. Rather than yield, the crop production and area are usually asked for in the interview.

5.1.1 Farmer recall and prediction

When crops have already been harvested at the time of the interview, the holder or farmer is requested to recall from memory the amount that was produced and the size of the harvested area. The enumerator conducts the interview either at the farmer's house or the enumerator, together with the farmer, visits each parcel or field that the holder cultivates, to determine the harvested area and recall the amount produced. If the harvested crop is still stored in granaries, the enumerator may visit the location with the holder. Visiting the parcels, fields or storage facility affords the enumerator the opportunity to verify the estimates provided by the farmer by physically observing the locations (Casley and Kumar, 1988). In developed countries such as the United States of America, rather than visiting the holder or farmer, telephone interviews are often used to obtain final harvested quantities and acres when the harvest period is over (Aune and Vogel, 2006).

Farmer prediction is the method in which the holder or farmer is asked to estimate the amount of production of the crop that has not been harvested. Farmer prediction provides a pre-harvest estimation, while farmer recall gives a post-harvest estimation.

In a typical agricultural survey, the sample farmers or holders would be asked to provide their expected quantity of crop that is still in the field, based on their past experience, by comparing current crop growth performance to that of previous years (David, 1978). Usually, the enumerator and holder would visit each parcel that the holder is cultivating to visually inspect the crops planted, for the holder to provide his or her estimate of the produce. Singh (2003) recommends that the holder be interviewed at the maximum growth level to improve the accuracy of the holder's estimate. However, in a general-purpose crop production survey in which many crops are included in the survey questionnaire, this may not be feasible because different crops have different cropping calendars.

The total harvested produce from all parcels cultivated by the sample holder planted to a crop is the total crop production. The crop yield estimate is then obtained by dividing the total crop production by the total harvested area. This is usually done by the enumerator. There are also surveys in which the yield at the holding level is not derived. Rather, the crop yields at the domain level, whether provincial or district, is computed by dividing the total crop production estimate by the total crop area harvested at that specific level. When this is done, appropriate survey weights should be incorporated into the estimation process. Usually, the farmer recall and farmer prediction methods measure economic yield.

When the holder provides the estimates in local units, the enumerator must convert them to standard units such as kg or MT per ha, using appropriate technical conversion factors.

Because the interview with holders do not require any instruments to measure production and area, this data collection method can be implemented in a larger sample. However, because this method depends, to a great extent, on memory recall or estimation based on previous experience, the resulting estimates may incur bias. For example, in many developing countries where illiteracy is prevalent in smallholder farmer populations, farmers may provide fictitious estimates, thereby undermining data quality (Kelly *et al.*, 1996). Depending on the prevailing situation, the holder may overestimate or underestimate his or her crop production. Holders who wish to convey a lack of government support may underestimate production, while those who are optimistic about the economy may overestimate. Some farmers use local units when measuring their production. They may also round up or down their total production. For example, when sacks are used for measuring and the total production includes a fraction of a sack, the farmer could omit this fraction or round it up to one bag. Other sources of bias in surveys using farmer declarations, that have been identified in many studies in different countries, include failure to quantify in-kind payments, non-standard harvest units, deliberate over- or underreporting, low accuracy with longer recall periods, historical average production factors, poor-quality responses in lengthy interviews, and insufficient supervision (David, 1978; Fermont and Benson, 2011; Sud *et al.*, 2015).

In general, information campaigns on upcoming surveys improve the survey response rate and the data quality provided by the respondents. In the case of a crop survey, an information campaign raises awareness among farmers and encourages them to participate. If they know the survey objectives and how the data they provide will be used, the farmers are more likely to provide true estimates.

Visual inspections by the holder and the enumerator improves farmers' estimates of production. In addition, omissions in production estimates can be reduced by including specific questions on in-kind gifts and payments (own-produce) made by the household. However, adding questions and requiring the farmer to visit all parcels he or she cultivates would result in a longer time being required to complete the survey questionnaire and may reduce the farmers' interest in cooperating. The survey questionnaire should therefore be streamlined to include only the important questions, so as to lessen respondent burden.

If local units are used to determine crop production, conversion factors must be established at the onset. These conversion factors are more effective at the district or provincial level, rather than at the national level.

The methodological study conducted by Verma *et al.* (1988) provided strong evidence in support of farmer declarations (recall and predictions) for estimating crop production, although not necessarily for crop yield. The study was undertaken in Benin, the Central African Republic, Kenya, Niger and Zimbabwe. It was designed such that farmer declarations could be compared with crop-cut estimates and with actual production based on the whole plot harvest method. The farmers' estimates were obtained twice: once two to four weeks before harvest and once after the tharvest. It was found that both farmer estimates were closer to the actual production and had a lower variability than the crop-cut estimates. While the farmer estimates were fairly accurate, the crop-cut estimates were significantly higher than the actual production estimates. Diskin (1997), however, cautioned that while the Verma study showed that farmer declarations are better than the production estimate from crop-cuts, farmer estimates of crop yield may not be as precise as their harvested area estimate may also present measurement errors.

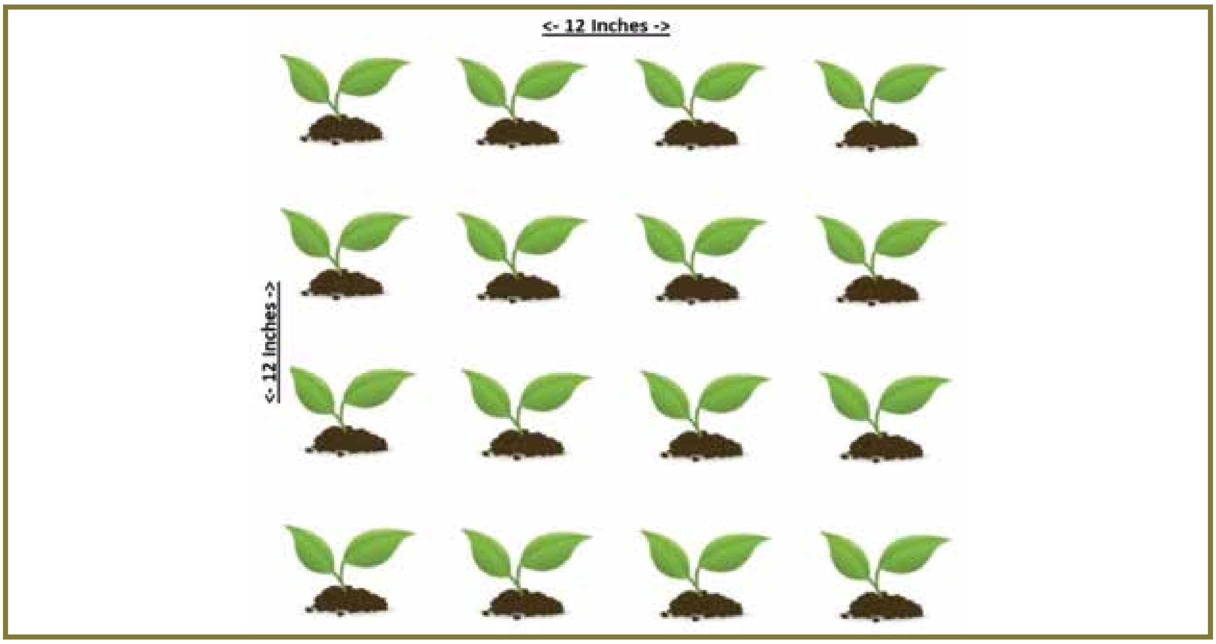
The results of the review, undertaken by Rozelle (1991), of six Cornell studies to examine farmers' abilities to provide yield estimates directly was not conclusive . Farmers in China and Indonesia provided precise yield estimates of their crops, while those in the Philippines and Nepal gave less reliable yield estimates. The farmers in Malawi found it difficult to give yield estimates because they could not adequately estimate the crop area planted.

5.1.2 Expert assessment

The visual assessment of crop performance by way of experts, such as field agronomists and extension officers with several years of experience, observing the conditions of the crop in the field regarding for example colour, plant vigour and plant density could form a basis for estimating crop production or yield. This type of procedure could be used to estimate biological yield and is usually referred to as eye assessment. Expert assessment entails combining eye assessments, field measurements and the application of empirical formulas for estimating crop yield. In developed countries such as the United States of America and Australia, this method is widely used because of the mechanized farming systems prevailing there, where crops are planted with equal and precise row spacing (Fermont and Benson, 2011). In developing countries, where smallholder farmers are predominant and, in most cases, crop planting is done with unequal row spacing, it would not be feasible to apply this method under the current farming practices of the traditional farming systems.

For instance, the crop yield for cereals can be estimated by assessing the various essential components and using the following formula: average number of grains per head multiplied by the average number of heads per 5-m row, and dividing that product by a constant K that is dependent on row spacing and grain weight. To make the average number of heads per 5-m row representative, the sample size of the number of counts within a field should not be less than 10 (DPI, 2010; Fermont and Benson, 2011). As the accuracy of the method primarily depends on the level of expertise of the personnel involved, caution should be applied when using it on large-scale samples (Fermont and Benson, 2011). Casley and Kumar (1988) reported that extension officers may be biased in favour of their own work, resulting in higher estimations, if they are engaged in estimation that relates to their own work area.

FIGURE 5.2. CROP PLANTED WITH EQUAL SPACING.



Source: https://gardeninminutes.com/square-foot-gardening-spacing.

The possibility of using a single team of experts throughout a study ensures consistency in yield estimation (Rozelle, 1991). Area estimation is considered as one of the potential sources of bias in yield estimation; however, the expert assessment method does not require the estimation of the area, thereby ruling out the likelihood of this source of bias. Expert assessment also has the comparative advantage of being suitable for large-scale investigations, compared to crop-cutting and farmer estimation methods.

The main disadvantage of this method is that it needs high-calibre experts that may not be readily available to go to specific locations. The use of different teams of experts may result in variations in the mode of estimation, which has the corresponding effect of increasing the error margin. In addition, expert assessment is perceived to be excessively subjective, as the accuracy of the yield estimation depends primarily on the level of available expertise compared to other methods such as crop cutting or whole-plot harvest, which must follow stringent general rules.

5.1.3 Crop cards

The crop cards method is intended to encourage farmers, particularly smallholder farmers (who constitute the highest percentage of the farmer population in most developing countries), to develop the habit of regularly recording their harvest quantities. This method is especially suitable for crops with staggered maturity, such as banana and cassava, or that require multiple pickings, such as beans and tomatoes (Fermont and Benson, 2011). Farmers find it difficult to remember or memorize their produce because harvests are done over time. Recording harvests using crop cards will help them to determine their total crop production for a given reference period. The crop card method typically measures economic yield.

The method is carried out by initially distributing a set of crop cards to sample farmers or farms by a crop card monitor (CCM). Each farmer is expected to record the quantity of harvest (usually in local units) after each harvest operation. The CCM pays regular visits to the sample holder to monitor the recordings, assist them in correcting any mistakes, and to motivate and encourage them to continue the process. The CCM collects all cards from the farmers at the end of the survey period for processing. The processing may include adding up all recorded quantities for each farmer to estimate total crop production or total harvested yield for individual farmers.

To ensure consistency and unambiguous recordings, the farmers or knowledgeable family members selected for this type of survey are usually given adequate training in appropriate measurement and recording techniques to compile the various sections of the crop cards (Fermont and Benson, 2011; Nelson and Swindale, 2013). The training of farmers is considered an essential aspect of the survey operation, because failure to make the selected farmers familiar with the crop cards may result in various recording complications. Unlike in developed countries, most developing countries are unable to conduct such training for farmers, mainly because of limited resources. It is, however, in developing countries that illiteracy among smallholders is more prevalent.

In mitigating the effects of variation in local unit measurements across locations, appropriate standardized measuring units should be introduced; for example, standard-size buckets or containers could be distributed to the selected farmers, with well-explained instructions as well as illustrations (Fermont and Benson, 2011; Sud *et al*., 2015).

A crop card method that is well-planned and properly implemented with adequate supervision can give accurate and reliable yield estimates of crops with or without extended harvest periods, compared to the farmer recall method. More efficient random sampling techniques in the selection of farmer respondents can be used in the crop card method than in the crop-cutting method.

The crop card method, however, could be affected by significant biases as already discussed, if care is not taken in its planning and implementation stages. As the method requires training selected farmer respondents, it is relatively more costly than farmer recall.

The Uganda Bureau of Statistics tested the crop card method in the 2005/2006 National Household Survey; however, it encountered several irregularities caused mainly by inadequate manpower for regular monitoring and certain other issues discussed above, such as farmer illiteracy, resulting in wrong recordings and variations in unit measurements (Ssekiboobo, 2007; Fermont and Benson, 2011). In comparing the crop card and farmer recall methods, some studies reported that crop card production estimates for certain crops such as cassava, banana, maize, beans and sweet potato were lower than the farmer recall production estimates (Sempungu, 2010; Carletto *et al*., 2010; Fermont and Benson, 2011). This indicates that farmers are either overestimating their production during interviews or underestimating the crop card records. In practice, if crop cards are recorded by literate farmers or family members and the recording is regularly monitored, it is expected that crop cards would provide more reliable results than farmer interviews, especially for crops with staggered harvests.

5.2 MEASURING YIELD THROUGH OBJECTIVE MEASUREMENTS

Objective crop measurement methods entail direct field crop measurements. These are widely used in both developed and developing countries, and are usually considered as reliable and objective means of obtaining crop production or yield estimates. Researchers generally employ these methods in experimental studies to compare productivity and efficiency. Objective crop measurement methods include crop cutting, whole-plot harvest and remote sensing.

5.2.1 Crop cutting

The crop cutting method was adopted by FAO in the 1950s as a standard method for estimating crop production or yield after its initial experimentation in India in the late 1940s (FAO, 1982; Murphy, Casley and Curry, 1991; Fermont and Benson, 2011). The method is considered the most widely used objective form for measuring crop yield because of its reliability, cost-effectiveness and less labour-intensive nature, compared to whole-plot harvest. The crop cutting method involves harvesting one or more randomly located subplots within one or more selected cultivated fields with a specified crop or crops and weighing the harvested crop(s) after drying. Typically, the harvesting is carried out by either the enumerator or the farmer. The crop-cutting method measures biological yield. Annex 6 presents more details on how to implement the crop-cutting method.

There are various ways of implementing this method, including locating a central plot within the selected field(s) or randomly locating one or more subplots within the selected field(s). The shape of the subplot could be a square, a rectangle, a triangle or even a circle (Fermont and Benson, 2011; Sud *et al*., 2015). The crop-cut method was first implemented by measuring the yield of one random subplot in a sample parcel. The crop is harvested by the data collector in the selected subplot, dried, processed and weighed. To measure variability in yield in a parcel, Fielding and Riley (1997) recommended that crop cut be implemented in two randomly selected large quadrants. Multiple small quadrants, selected systematically, were employed in estimating the cereal yield in Botswana by Norman *et al*. (1995). Intercropping is measured by recording the data of each crop separately for each sample quadrant. Rozelle (1991) recommended a less labour-intensive approach: the "five-point" method, in which the data collector harvests five 1-m^2 quadrants located in the corners and at the centre of a plot.

Countries with the requisite resources and the desire to increase the level of precision would have at least three subplots randomly located within each selected field. The inherent variability that exists within farmers' fields is captured by multiple subplot crop cuts (Nelson and Swindale, 2013).

For example, crop yield per unit area or subplot is calculated as the total crop production of the subplot (kg/m^2) divided by the total harvested area or subplot (ha). Thus, total crop production is obtained by multiplying the crop yield per unit area or subplot by the total planted area or field. Crop production is measured in MT while crop yield is measured in MT/ha.

For data collections having widespread coverage, choosing holdings to apply the crop-cutting method is more cost-effective when two-stage sampling or cluster sampling are used, if the PSUs of the cluster sampling, which could be a village or a group of segments (for the area frame), are selected from which holdings will be drawn. This sampling technique ensures that the data collectors will have a balanced workload and that only the holdings in the selected PSUs need to be listed for the second stage of the selection process. However, the resulting estimates may not be as precise as those obtainable through SRS.

Knowing the shape of the subplot (for example a square, rectangle, triangle or circle) before the experiment is carried out eliminates any errors associated with estimating crop yield per unit area or subplot and the ensuring final yield estimation (Poate, 1988; Sud *et al*., 2015). If due diligence is followed and correct procedures are observed, the difference between the crop-cutting yield estimate and the whole-plot harvest yield estimate will not be statistically significant (Rogers and Murfield, 1965; Sud *et al*., 2015). Comparatively, a crop cut of a reasonable subplot size is less costly than a whole plot harvest of the same field.

Crop cutting is perceived to be a reliable and objective measure of crop yield. However, several studies have revealed that the method could be affected by significant biases if care is not taken at the planning and implementation stages. Murphy *et al*. (1991) highlighted the following sources of error incurred in experiments on the estimation of agricultural production in Africa in a study presented byScott,Marchant andVerma at the Forty-seventh session of the International Statistical Institute in 1989:

Selection bias. Data collectors tend to select locations near the centre of the plot. Also, even well-trained data collectors may decide to move the sample subplot if the random coordinates obtained fall on less fertile area. When a part of the random coordinates falls outside the parcel border, the data collector usually moves the whole sample subplot inside the parcel border. These examples distort the selection probability of subplots that is used in the computation of yield.

Coverage error. When the parcel has an irregular shape, the area near the uneven contour may not be considered when selecting for the subplot. When the sample subplot border falls on only part of a plant, data collectors tend to include the plant rather than to exclude it.

Other measurement error. Because the sample subplots are small in size compared to the parcels or fields, data collectors would be likely to harvest the crop-cut from the subplots thoroughly, compared to the real-life situation in which grains are left on the ground. The subplot yield is an estimate of biological yield, as it covers the total harvest from plants rather than what was produced by farmers for sale or consumption. Another type of error occurs when measuring the lengths and bearings of an irregularly shaped parcel: the data collector may tend to approximate the uneven contour with straight lines and in so doing, incur errors in area measurement.

The crop-cutting method is costly, labour-intensive and time-consuming compared to the farmer declaration method. The interest of most farm holders is to obtain an estimate of the economic yield (that is, yield after subtracting both harvest and post-harvest losses). However, the crop-cutting method measures the biological yield, which may not be as meaningful to the farm holders.

5.2.2 Whole-plot harvest

This method is regarded as the standard for crop yield estimation, because if appropriate procedures are followed, it does not incur any bias. The implementation of this method varies with the two types of crop maturity – those with well-defined maturity, such as cereals, and those with staggered maturity, such as banana and root crops. In the case of the former, harvesting can be conducted in a single operation, while the latter requires multiple harvests.

Before the harvesting operation is carried out, the data collector, together with the holder, must delineate all of the boundaries of the planted area to ascertain the correct total size of the farm or harvested area. To control any bias in the measurement of the harvested plot, appropriate steps, such as carefully approximating curved lines with straight lines, should be considered, as in most cases planted fields or plots are irregularly shaped (Fermont and Benson, 2011). The crop is then harvested with the data collector either observing the process or taking part in the harvesting. If the crop is a cereal, the grains are separated from the stalks and winnowed. After drying and weighing the harvested crop from the farm, the crop yield is obtained by dividing the total harvested crop weighed by the total harvested area. The whole-plot harvest measures the harvested yield.

If the crop requires multiple harvests, the data collector must participate in all harvests. The harvest is dried, if appropriate, and weighed. The weight for each harvest is recorded. The total weight will be the sum of the weights of all harvests for the cropping season or for the year.

The sources of bias, such as selection bias (for example, border effect, within-field variability effects, non-random location of subplots) and coverage errors that are associated with crop-cutting methods, are nonexistent when whole-plot harvest is conducted. Accordingly, whole-plot harvest is considered an almost bias-free method. Bias may occur only when irregularly shaped parcels are not measured well or when the weighing is not properly conducted.

Whole-plot harvest is the standard used to evaluate other methods for estimating yield. However, because this method requires the presence of the data collector during the harvests, it is perceived to be laborious and costly. Therefore, only a few plots can be measured thoroughly, given the limited resources for agricultural statistics. Therefore, this method can only be implemented in small-scale methodological studies and not in large-scale data collection efforts.

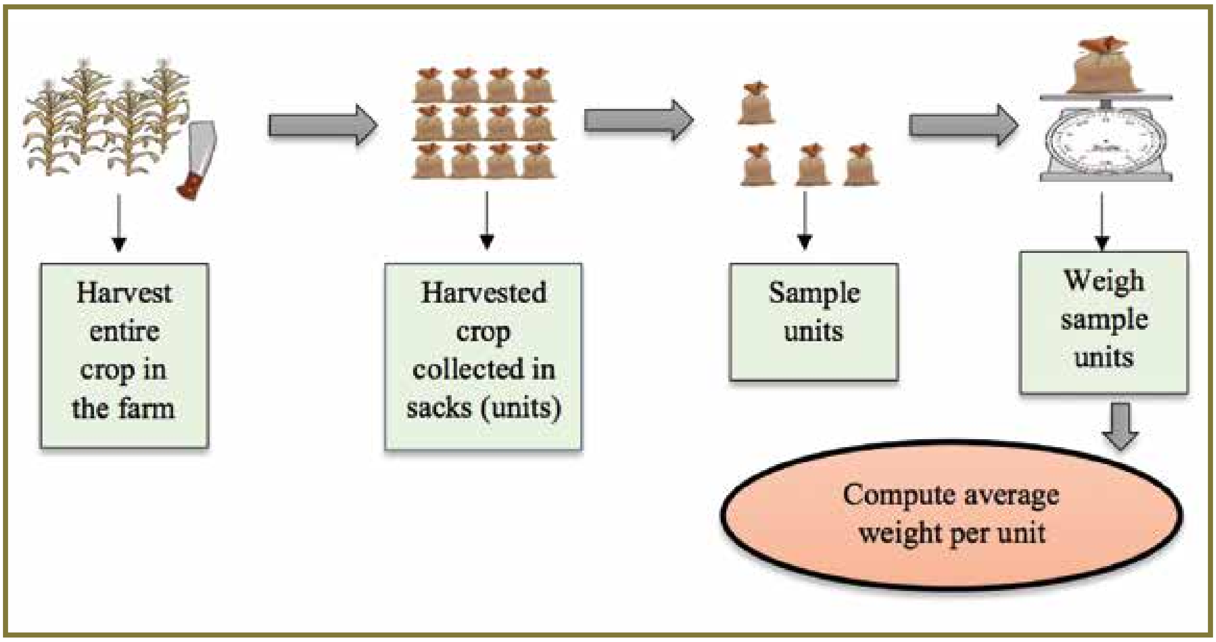
5.2.3 Sampling the harvesting units

This method is an offshoot of the whole-plot harvest method, that was developed to reduce the time required to weigh the harvest. According to this method, the data collector and the sample holder will also determine the crop area before the harvest begins. The data collector need not be present during harvest; however, he or she must convince the holder to store the harvested crop in storage units having uniform capacities. After the crop is harvested and processed, they should be stored in similar storage units of equal size, such as sacks, baskets or bundles. The data collector then randomly samples some of the filled storage units and weighs them, to obtain an average weight per storage unit. As the sample is usually obtained immediately after harvesting or before storage in granaries, the measurement of the average weight per unit includes some of the moisture content of the harvested crop (Casley and Kumar, 1988). Depending on post-harvest operations that are considered before the enumerator weighs the sample units, the method may render either the harvested yield or the economic yield (Fermont and Benson, 2011).

The total production or total harvest produced is obtained by multiplying the average weight of a storage unit by the total number of storage units counted from the harvest or storage facility. Yield will be derived by dividing the estimate of total production with the total harvested area.

Poate and Casley (1985) suggested that the method is well suited to measuring crop production at the holding level rather than crop production at individual plot level, as the holding may consist of more than one parcel or plot planted to the same crop. When the parcels are harvested in the same period and the harvested units are kept in the same storage facility, it is difficult to ascertain from which parcel they come from and, therefore, the parcel production estimate may not be as accurate. When the enumerator fails to visit the farm immediately after harvesting to make the necessary measurements, the holder may have already sold or consumed some of the produce. Rozelle (1991) proposed that in such situations, the data collector should ask the farmer for information on the quantities of harvests that have been previously used for domestic consumption, sold, given in-kind to friends or used as in-kind payments, for the period between harvesting and the time of the enumerator's visit to take the measurements.

FIGURE 5.3. ILLUSTRATION OF DECLARATION WITH HARVESTING UNITS.



Source: Poate and Casley, 1985.

When sampled farmers are not given information on using uniform storage units, they may not store the harvested crops as required, which may make it difficult to obtain the average weight per unit. For instance, maize can be stored in granaries without necessarily having to keep them in sacks.

As the farmer is expected to harvest the crop in a single operation and keep the harvested crop in uniform storage units immediately after harvesting, this method is not suitable for crops with staggered maturity and extended harvest periods that require multiple harvests over a longer and variable period (Fermont and Benson, 2011). However, the method is cost-effective for crops with well-defined maturities.

5.3 OTHER METHODS FOR ESTIMATING YIELD AND PRODUCTION

5.3.1 Administrative records

Administrative records can be used to validate the production estimates of certain crops and, in some cases, when other data sources are not available, they can also be the primary source. For example, given that the entire cotton crop will be ginned, the total production of cotton can also be estimated from the reports of all cotton gins. The purchase records from agrochemical industries could serve as another valuable source of information for estimating cash crops such as cocoa, coffee, shea nut, cotton, tea and sugarcane. These industries usually have well-organized procurement procedures such that data on the quantity purchased for each crop and season are always recorded. If the industries can share the summary of these records with the relevant NSO, then they can be used to estimate production.

For example, in Ghana, yearly production records for crops such as cocoa, coffee and shea nut are usually obtained from agrochemical industries, which are also known as Licensed Buying Companies (LBCs) (Ghana Cocoa Board, 2014). Linking these purchases to individual farmers is not difficult because proper record-keeping systems are in place. Similar well-organized systems have been reported in Sweden and Norway for sugar beet production (Bradbury, 1996a) and in Uganda for cotton production (Fermont and Benson, 2011).

The primary use of administrative records is to validate the production or yield estimates of crops. This method may not be effective when crops have varied uses; therefore, many industries need to be monitored. Also, this method cannot be applied when the crops in question constitute a staple food and are used for consumption by the household or livestock (Sapkota *et al*., 2016).

5.3.2 Crop modelling

The relationship between crop yield and crop variety, agrometeorological factors such as rainfall and soil condition can be modelled using statistical methods if a data set with sufficient number of observations is available. The resulting model can then be used for predicting yield given certain conditions. If an estimate of crop area is also available, then crop production can also be derived. This method measures biological yield and is usually applied only at a higher level of disaggregation, such as at the district or provincial level (Fermont and Benson, 2011). Crop yield can also be estimated on the basis of more complex crop growth models that consider crop physiology. The International Rice Research Institute, for example, maintains a crop growth simulation model that uses real-time and historical weather data such as solar radiation, minimum and maximum temperature, daily average wind speed and rice leaf growth rate (Setiyono *et al*., 2014).

Establishing crop models requires a large empirical data set and expertise, elements that may not be available in the NSOs of developing countries. The crop models developed in some countries may not be applicable in others because of the countries' inherent differences with regard to the relevant environmental conditions.

5.3.3 Allometric models

An allometric model defines a mathematical relationship between plants' morphological characteristics and crop yield (Sapkota, 2016). Using the model, the morphological characteristics of a selected number of plants can be measured to predict the biological yield. The model is usually based on plant characteristics that can easily be observed and measured, preferably without destroying the plants. For example, the plant height and the ear length of corn can be used to predict the biological yield (Fermont and Benson, 2011).

Very few studies investigate the use of allometric models in predicting yields for large areas of interest, such as a province or the entire country. The accuracy of this method has not been examined well, compared to the other methods discussed in this chapter.

5.3.4 Remote sensing

Remote sensing is a new procedure for estimating crop production or yield (biological yield) that is being tested in several countries, including the United States of America, China and India, to maximize its potentials (Zhao *et al*., 2007). The principle behind the remote sensing procedure is that green plants possess a unique feature called a spectral signature, which can determine the state, structure and composition of the plant by using satellite imagery techniques. These images captured by satellite are data (spectral data) that are used to construct vegetation indices such as the normalized difference vegetation index (NDVI) and the ratio vegetation index. To properly estimate crop yield, it is necessary to conduct fieldwork to accurately establish the association between the NDVI values and crop characteristics, such as crop types and yield (Fermont and Benson, 2011).

The use of remote sensing in forecasting crop yield is still in the experimental stage. A project of the Asian Development Bank in support of GSARS is exploring the well-established relationship between backscatter and rice growth stage to estimate the rice crop area. Backscatter is the echo from the ground of the microwave signal emitted by a satellite aperture radar (SAR) instrument. This echo is also recorded in the satellite image. When newly planted, the backscatter from rice is weak and the SAR data image becomes dark. As the rice plant grows, the backscatter becomes stronger and the brightness of the image increases. The yield is then determined using information gathered from the ground (such as the rice variety planted and historical yield data) and the relationship between biomass and SAR backscatter.

The Mahalanobis National Crop Forecast Centre generates remote-sensing-based forecasts on the area and production of selected crops and provides monthly assessments of the drought and flood situation in India (Shankar Ray *et al*., 2014). The forecasting method employs both radar and optical data to provide a 90-percent area coverage at an accuracy of 90 percent. The methodology uses both microwave and optical data and is designed to provide an accuracy of 90 percent. To forecast crop area, square grids of 5 km x 5 km are classified according to the proportion of crop in the square grid. Twenty percent of the square grids in each of the four strata based on crop proportions (>75 percent, 50–75 percent, 25–50 percent and <25 percent) computed from past data are sampled. The remote sensing data of the sampled grids are classified into crop types using satellite data from different observation dates. Satellite images of the same area are captured every 20–25 days. To estimate the crop yield, econometric and agrometeorological models have been developed. The crop production is then determined from the forecasted crop area using remote sensing and the yield from the crop models.

The International Rice Research Institute (IRR) forecasts rice yields based on a crop growth simulation model using real-time and historical weather data. Based on its pilot study conducted in South and Southeast Asian countries, the incorporation of remote-sensing data (SAR) into process-based crop models improves the yield estimation for actual yields, because the remote sensing data better captures the variation in the environmental conditions of a large area (Seyitono *et al*., 2014).

Notwithstanding the viability of the remote sensing method, several studies have suggested conducting further research to address certain basic challenges, including developing high-resolution satellite imagery equipment that can accurately capture crop images (that is, in smallholder fields, intercropping fields and crops with varied dates) and cloud coverage (Fermont and Benson, 2011; Nelson and Swindale, 2013).

The use of remote sensing in the estimation of crop area, yield and production can supplement the current data collection and estimation practice. However, it requires staff skills that exceed the technical capacity typically possessed by staff, access to satellite imagery, GIS and high-speed computers, which may not be readily available in the NSOs of developing countries.

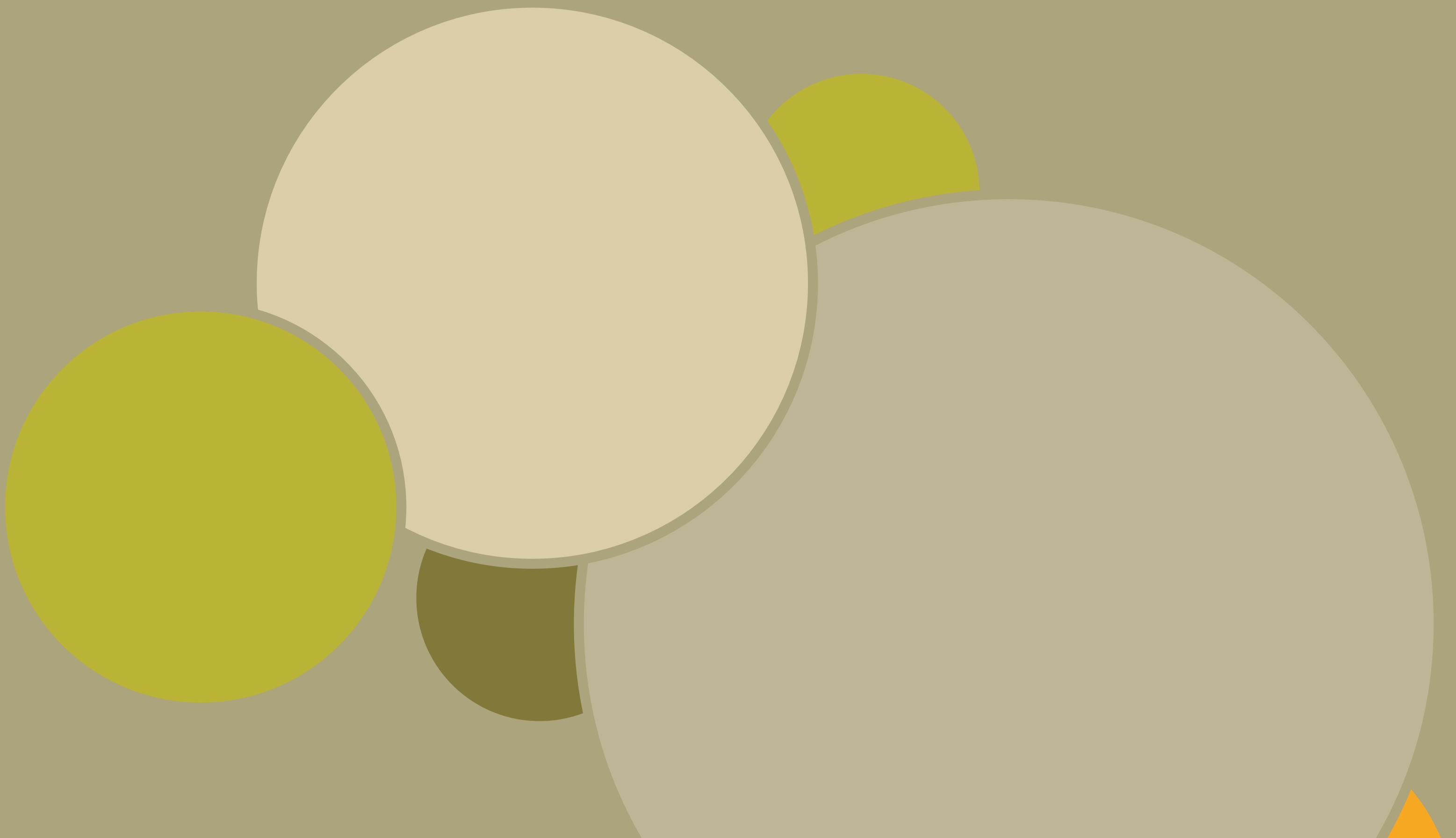
5.4 COMPARISON OF EXISTING METHODS FOR MEASURING CROP PRODUCTION OR YIELD

The cost-effectiveness, level of estimation and precision of each method discussed in this chapter are summarized in table 5.1, which was adapted from Sapkota (2016). The third column of the table, titled "Scale", provides the level of estimation at which the method can be used. "Landscape" means that the method can be applied to a large area of interest, such as a district or a province.

TABLE 5.1. COMPARISON OF EXISTING METHODS FOR MEASURING CROP PRODUCTION OR YIELD.

Method	Cost-effectiveness	Scale	Precision in estimation, errors and biases
Farmer's estimate	Cheap and fast method that saves time and money	Farm to landscape	Fairly accurate estimation; however, needs adequate supervision. Subjective. Farmers may deliberately overestimate or underestimate.
Expert assessment	Moderately cost-effective	From farm to landscape level	Chances of error increases if different teams of experts are used or extension people are used to estimate yield in their own area. Subjective.
Crop cards	Cost- and labour-intensive	Field to farm level	Bias because of illiteracy, use of local units
Crop cut	Time- and labour-intensive	Field, farm and sometimes landscape level	Tendency to overestimate
Whole plot harvest	Cost- and labour-intensive	Plot level, farm level, case study	Almost bias/error-free
Sampling harvest unit	Cost-effective	Farm to landscape	Error-prone in the condition where farmers harvest from multiple areas at one time; cannot be used with staggered harvesting
Purchaser's insurance record	Cost-effective	Field scale	Suitable for cash crops only with no household consumption
Crop modelling	Cost-effective	Landscape	Less, if adequately parameterized and calibrated. Does not include induced improvements in agricultural tecÚology.
Allometric models	Cost-effective	Field scale	Suitable for few crops only
Remote sensing	Cost-effective	Landscape	Chance of error in cases where different crops have the same signature

Source: adapted from Sapkota (2016).



6

Area and yield estimation under mixed cropping

With limited land available for cultivation, smallholders in developing countries usually prefer to plant a combination of crops, instead of only a single crop. With the appropriate combination of crops in the field, soil nutrients can be balanced without the aid of commercial fertilizers and weeds, insects and plant diseases can be minimized. For these reasons, mixed cropping is likely to lead to efficient land use.

Studies have shown that mixed cropping is prevalent in many countries. In India, because mixed cropping is dominant, the manual on crop production statistics of the Central Statistics Organization has specific data collection and estimation procedures for mixed crops. Microstudies conducted in Rwanda, Burkina Faso, Senegal and Niger have found that mixed cropping is practiced in each country (Kelly, 1996). Casley and Lury (1981) reported that in Ghana, 84 percent of the area under seasonal crops contained a mixture of crops, and that in Botswana, 90 percent of the area under millet and more than two-thirds of the area under sorghum contained other crops. Figures from the 1990–1991 agricultural census in Uganda indicate that, with 80 to 90 percent of their area being in mixed stands, intercropping may have become more important for maize, beans, millet and groundnuts, whereas the intercropping of cassava remained at 50 percent (Fermont and Benson, 2011).

Mixed cropping involves two or more different temporary or permanent crops grown in the same field. An intercrop is a crop that is planted between rows of another crop, while associated crops refer to the combination of temporary and permanent crops planted in the same field (FAO, 1982). Because of these characteristics, determining crop yields and calculating the harvested or planted areas of mixed crops to analyse their productivity is not straightforward. The number of crops planted in the field may vary. As the number of crops increases, the complexity of estimating each crop area also increases. The vegetative cycles of the crops in the field and, consequently, their periods of harvesting, may also differ. Collecting the correct production data for each crop and attributing the appropriate crop areas requires due diligence and resources.

One challenging issue arising in survey operations, especially for a general-purpose crop survey, is how to record the crop areas in mixed fields. A simple approach would be to record the total area of the mixed field together with the mixed crop combination. Deriving the total crop areas based on this approach, however, can be problematic when there are many different combinations of mixed crops, making the summary tables intractable.

Countries use different approaches in recording the production and area of mixed crops. The same harvested area could be used to determine the yields of two or more mixed crops. If this happens across all sampled farms with mixed crops, then the resulting estimate for the total agricultural area is likely to be an overestimate, and the yield of the crops in mixed crop combinations would be underestimated. To avoid overestimation of the agricultural area, on the other hand, some countries only report the areas for the principal crop in mixed fields, discarding the data of the secondary crops. This could also lead to underestimation of production and yields. If appropriate methods for determining crop areas and yields in mixed crops are not implemented, because mixed copping is prevalent in many developing countries, the estimations of production and yield at the national level could be biased. Kelly (1996) found that failure to account for all crops in mixed fields could result in underestimation of productivity and income.

FIGURE 6.1. EXAMPLES OF MIXED STAND.

Mixed stand of cassava and plantain

Mixed stand of rice and maize

Sources: ENSEA/WAAPP agricultural surveys, 2013–2014 and 2015–2016.

Cashew nut tree in maize field

Sources: ENSEA/WAAPP agricultural surveys, 2013–2014, 2015–2016

Mixture of maize and watermelon

Global Strategy cost of production field test, March 2018, Botswana

This chapter presents methods for apportioning areas to crops planted in mixed fields and four strategies for estimating production and yield in situation of mixed crops. In this chapter, the term “mixed cropping” can be used interchangeably for mixed crops, intercrops and associated crops, unless otherwise specified, as many of the methods for deriving areas and yield statistics apply to these three types of mixed cropping.

6.1 ESTIMATING CROP AREAS UNDER MIXED CROPPING

This section presents various methods for attributing areas to mixed crops. The methods may be classified in two categories: allocated area and imputed area.

The area allocated to a crop is the fraction of the physical field area to which the particular crop has been allocated. The sum of the allocated areas of the different crops in a mixed field should always be equal to the total area of the field.

Imputed area, on the other hand, is the area that would have been occupied by the crop had it been cultivated in pure stand (FAO, 1982). Unlike the allocated areas, the total imputed areas of crops in a mixed field are not expected to be equal to the field’s total physical area.

Some criteria to decide whether to use an apportioning method or not are provided in annex 8 to this Handbook.

6.1.1 Methods for determining allocated area

Various methods of crop area allocation are practiced in developing countries. The simplest method divides the total field area sown *equally* among the component crops. Although easy to implement, it results in biased estimates of the crop areas. Therefore, when these areas are used with yield data, biased crop production estimates will result.

Countries may also practice allocating each component crop based on each crop’s total area sown, that is, the allocated area for crop *A* is equal to the total area sown for crop *A*. However, this method fails to reflect the damages encountered by the crops during the seeding time, thus leading to overestimated crop areas and underestimated yield estimates. These methods are obviously not appropriate for mixtures that have varying crop proportions, as they do not reflect the relative importance of each crop in the parcel.

While the practice of sowing mixed crops in the same field is relatively common in developing countries, the methods of sowing are not uniform. At least two cultivation methods are currently practiced. The crops in mixture are sown either row-wise separately or mixed altogether. The sowing method and the types of crops, whether all temporary, all permanent or associated crops, may influence the choice of following allocation methods.

Temporary crops: apportioning using plant density

This method apportions the whole field among the component crops according to the plant density ratio. A number of plots are first randomly selected, and the number of plants in said plots are then counted. To calculate using the plant density ratio, let i be a component of a crop mixture with total area A. Also, let n_i be the average number of plants of component crop i per unit area and N_i be the number of plants per unit area of component crop i in pure crop stand (Sud *et al.*, 2015). Data on the number of plants per unit area under pure stand can be derived using objective measurement techniques. Plots planted to crops under pure stand that are randomly sampled can be observed. The averages of these data can then be considered as the corresponding data for pure stand. Under this method, the area allocated for a specific component crop i is *area of crop* $i = A \times \frac{n_i/N_i}{\sum_j n_j/N_j}$.

For illustration purposes, suppose that there are three crops in the crop mixture over an area of 0.8 ha. Suppose also that there are 100 plants of *crop A* when sown in mixture, and 2 500 plants when *crop A* is sown as pure; 18 plants of *crop B* when sown in mixture, and 25 plants when sown as pure; 80 plants of *crop C* when sown in mixture, and 200 plants when *crop C* is sown as pure.

Then, the plant density ratio is computed as

$$\text{plant density ratio} = \frac{\text{no. of plants, crop A mixed}}{\text{no. of plants, crop A pure}} : \frac{\text{no. of plants, crop B mixed}}{\text{no. of plants, crop B pure}} : \frac{\text{no. of plants, crop C mixed}}{\text{no. of plants, crop C pure}}$$

$$= \frac{100}{2500} : \frac{18}{25} : \frac{80}{200} = .04 : 0.72 : 0.4$$

Thus, under this method, the apportionment of the whole area is as follows:

$$\text{area of crop A in the crop mixture} = \frac{0.04}{0.04 + 0.72 + 0.4} = \frac{0.04}{1.16} \times 0.8 = 0.0276 \text{ ha.}$$

$$\text{area of crop B in the crop mixture} = \frac{0.72}{1.16} \times 0.8 = 0.4965 \text{ ha}$$

$$\text{area of crop C in the crop mixture} = \frac{0.4}{1.16} \times 0.8 = 0.2759 \text{ ha}$$

As this is an objective method carried out on the basis of physical observations, precise estimates are expected to be obtained. This method is especially helpful when the component crops are sown as mixtures through narrow row spacing or broadcast sowing, when no row arrangements are defined.

In Rwanda, where intercropping is widely practiced, the Crop Assessment Report (MINAGRI, 2009) stated that the density of each crop in mixed parcels are roughly estimated *de visu* and standardized by dividing the density of a particular crop by the sum of the total densities of all crops in the field. The allocated area for a specific crop is then derived as a product of this ratio and the total physical area of the field. This approach does not require information regarding the densities of crops under pure stand. It is also used to derive estimates of the percentage of land allocated to specific crops at a higher level of disaggregation, for example at the district level.

Temporary crops: apportioning using row ratio (row intercropping)

Row ratios can also serve as a basis for apportioning the area under each component crop, especially if the crops are sown in separate rows. In this method, three places are chosen at random in the selected field. At each place, the average number of rows is determined by counting the number of rows in a specified length. The ratio of crop area is the comparison of quotients, denoted by q_i, when the number of rows of a component crop i in a mixture in a specified length is divided by the number of rows of i as per normal spacing in pure stand in a specified length (Sud *et al*., 2015). Using this method, if a row intercrop covers an area A, then

$$\text{area of crop } i = A \times \frac{q_i}{\sum_i q_j}.$$

For example, suppose that crops A and B are intercropped in an area of 0.8 ha. Let the numbers of rows of A and B in a specified length in intercropping be 32 and 1, respectively. Also, let the number of rows of A and B in a specified length in pure stand be 40 and 5, respectively.

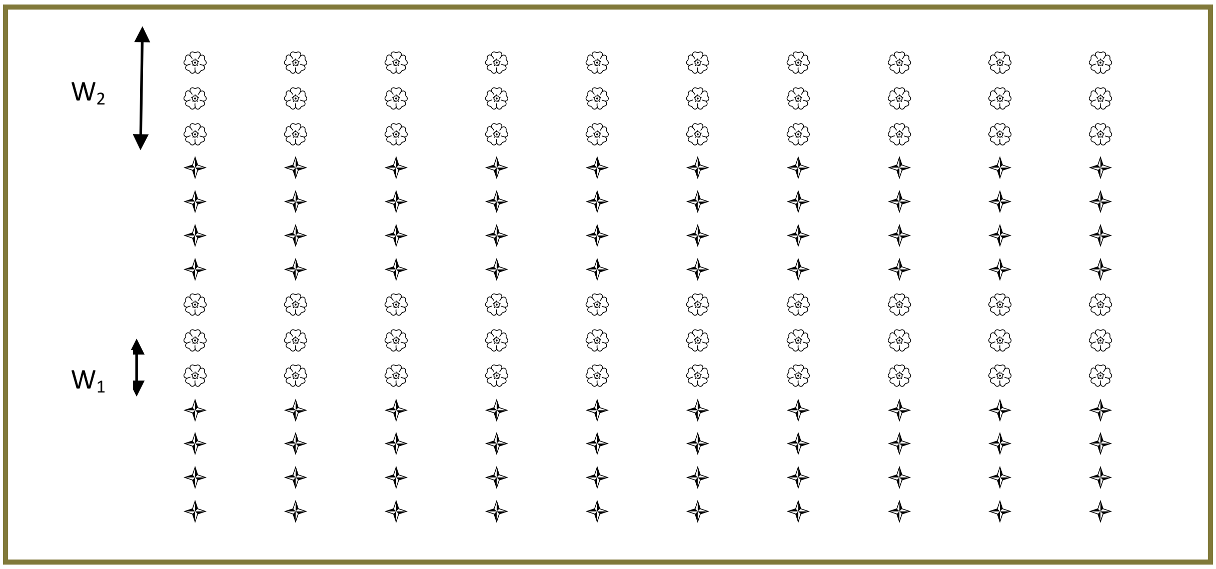
Therefore, the *crop-row ratio* $= \frac{32}{40} : \frac{1}{5} = 0.8 : 0.2$

In reality, the areas of component *crops A* and *B* in the intercropped land are 0.64 ha and 0.16 ha, respectively.

Temporary crops: apportioning using width ratio (strip cropping)

When component crops are planted in separate distinct groups of rows or strips as illustrated below, the average widths of the strips can be used to allocate the area among the intercropped crops. Similar to use of the row ratio, three points in the field are randomly selected and the average width of the component crop in the strip intercropping field is determined. This will serve as the basis for apportioning the intercropped field among the component crops (Sud *et al*., 2015).

FIGURE 6.2. ILLUSTRATION: APPORTIONING OF AREA BASED ON WIDTH RATIO.



Source: Sud *et al.*, 2015.

In the illustration above, suppose that crop ❁ is intercropped with crop ✧ in strips. To apportion the crop area between the two crops, the ratio of the average widths of the strips can be used. Let W_1 and W_2 be the average widths of strips of crops ✧ and ❁, respectively.

Then, the area for crop ✧ will be $W_1 / (W_1 + W_2)$ *x total area of the mixed fields*.

For example, suppose that *crops A* and *B* are in a strip intercropping pattern. The total area of the intercropped field is 1.0 ha. Some strips of the crops are randomly selected and measured. If the average widths of strips of component *crops A* and *B* are 2 m and 3 m respectively, then the areas planted to *crops A* and *B* can be allocated such that

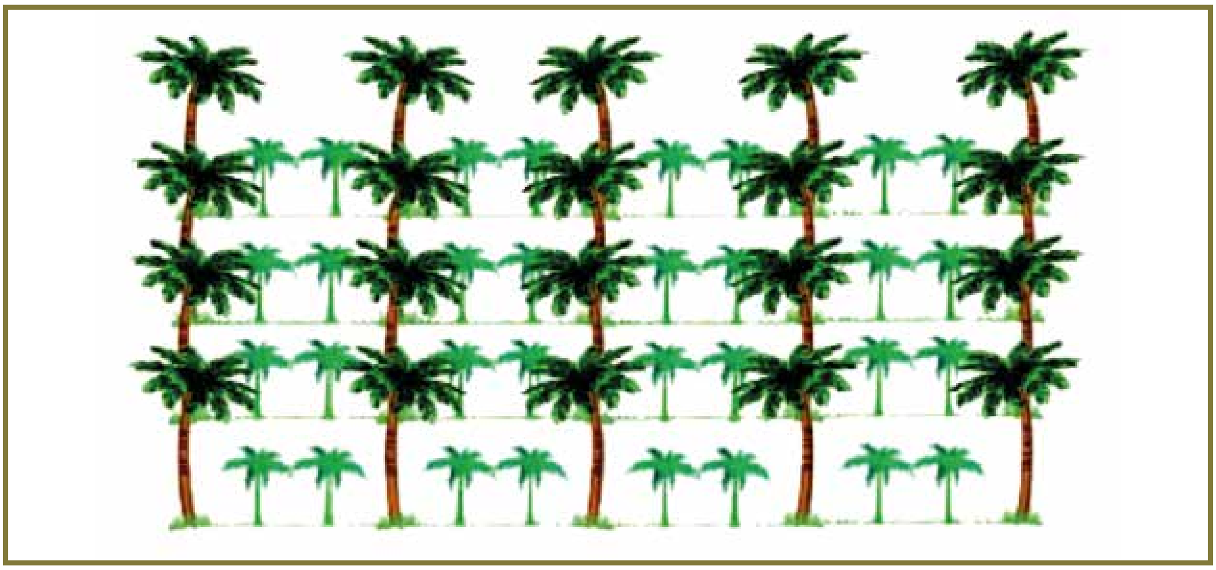
$$area\ of\ crop\ A = \frac{2}{2+3} x1 = 0.4\ ha, \text{ and}$$

$$area\ of\ crop\ B = \frac{3}{2+3} x1 = 0.6\ ha$$

Permanent crops: apportioning area based on tree counts

In the case of mixtures of two or more permanent crops, the area may be apportioned based on the number of trees of each component crop by adjusting the area to the pure stand (Sud *et al*., 2015). The number of trees and plants of each permanent crop may be recorded separately.

FIGURE 6.3. ILLUSTRATION OF A FARM WITH TWO PERMANENT CROPS: COCONUTS AND CACAO.



Source: Sud *et al.*, 2015.

Associated crops: apportioning area methods

As permanent crops are only planted once and harvested when ripened, these crops remain in the field for more than one year and occupy the whole area. Effectively, the orchard's entire area may be allotted to the permanent crops. However, this approach may lead to an excessive total cultivated area as temporary crops were also grown in the orchard. Therefore, the area of the orchard must be allocated between the permanent and temporary crops.

The area of the orchard allotted to permanent crops can be estimated using the canopy area of sampled trees. According to this method, the average canopy area per tree (πr^2) is estimated on the basis of randomly selected trees in the orchard. Multiplying the average canopy area per tree by the total number of trees in the orchard gives the estimated total area occupied by the permanent crops. The total area is then subtracted from the actual total area of the orchard to obtain the remaining area to be allocated among the temporary crops grown in the orchard. The apportioning among the temporary crops can be performed using any one of the methods previously described.

6.1.2 Methods for determining imputed area

The imputed area of a component crop is the area that would have been occupied by the crop had it been cultivated in pure stand. Thus, the imputed area measures the efficiency of mixed cropping relative to single cropping. In literature, this is also referred to as the land equivalent ratio (LER). Imputed crop area is used to derive the production of a crop from yield that has been measured either through objective methods (such as crop cutting), historical yield data or eye estimation. The imputed crop area is multiplied by the available data on yield to estimate production.

Suppose that A is the physical area of the field; c_i is the numerical value of the characteristic for crop i under the conditions of mixed cropping; and C_i is the value of the same characteristic for crop i under the conditions of pure stand. Then, the imputed area A_i of crop i is given as $A_i = (c_i / C_i) \times A$. According to FAO (1982), possible main characteristics that can be used for the imputation of areas are:

1. amount of seeds;
2. density of the plants, such as mounds or hills;
3. volume of production; and
4. commercial value of the produce.

The choice of the characteristic to be used for the computation depends on its relevance to the objectives of the survey and on data availability. Data can be obtained through farmer interview. As for the commercial value of the produce, data can be obtained from other sources such as village elders, agricultural extension workers or nearby markets. The plant density of mixed fields can also be estimated by an expert. For the imputation method to work, there should be a specific standard procedure for collecting and recording the additional data requirements. Another issue that has to be addressed is how to measure the characteristic for pure stand. Single-crop densities, yields and seed rates vary considerably across years and regions (Poate and Casley, 1985). Historical data on crops under pure stand can be studied to derive the corresponding standards.

Note that the areas allocated to component crops can also be derived on by rescaling the imputed areas of crops such that the allocated area A_i' of crop i is given as

$$A_i' = \frac{c_i / C_i}{\sum_i c_i / C_i} A = \frac{A_i}{\sum_i A_i} A$$

Example: using seed rate

When component crops are sown as mixtures without any row arrangement, the area under each component crop in the mixture may also be apportioned on the basis of adjusted seed rate. In this method, n_i is the quantity of seeds used to sow component crop i in the mixture and N_i is the normal seed rate of component crop i when sown as pure.

The normal seed rates are determined by taking the average of the seed rate information in pure stand from randomly selected sample fields. Using the imputed area method described above, the imputed area for crop A will be $(n_A / N_A) \times$ *Area of the field*. The imputed area for *crop B* will be similarly defined. The allocated areas for *crops A* and *B* will be the rescaled imputed areas.

To illustrate, consider an area under crop mixture of 0.4 ha. The quantity of seeds used to sow *crop A* and *B* from their mixture are 50 kg and 1 kg, respectively. Also, suppose that for respective pure stands of *crops A* and *B*, the seed rates are 120 kg/ha and 5 kg/ha. Then, the imputed area for *crop A* will be $(50/120) \times 0.4 = 0.16667$ *ha*. and for *crop B*, $(1/5) \times 0.4 = 0.08$ *ha*.

The imputed areas can be rescaled to derive the allocated area, such that the allocated area for *crop A* will be *[0.16667/(0. 16667 + 0.08)] x 0.4 = [0.16667/0.24667] x 0.4 = 0.2702 ha*. Similarly, the allocated area for *crop B* will be *[0.08/0.24667] x 0.4 = 0.13 ha*.

Apportioning of area among component crops under this method is influenced by several factors. When only the seeds – and not the population of plants in the field – are considered, cases in which the seed did not fully germinate or the survival of plants is not optimum could result in overestimation of crop areas. Also, the reporting of the seed rate by the farmer may not be as accurate, a common outcome in subjective methods. In addition, the size of the seed (or test weight) may influence the area of each component crop, which may result in an incorrect apportioning of crop area.

Example: using volume of production
For example, consider a field 2 ha in size that contains four component crops, *A*, *B*, *C*, and *D*, that have produce as 100 kg, 400 kg, 300 kg and 250 kg, respectively. Suppose that the average production/ha of *crops A*, *B*, *C*, and *D*, when cultivated in pure stand, are 500 kg, 1 000 kg, 750 kg and 1 000 kg, respectively. Then, the imputed areas based on the production volume of the different crops are 0.4 ha, 0.8 ha, 0.8 ha and 0.5 ha. Observe that the total imputed area is 2.5 ha, which is different from the total physical area. This implies that a 2-ha field with the mixed *crops A*, *B*, *C* and *D* yields the same overall production as a 2.5-ha field of the same crops in the same proportion in pure stand.

In the example above, to obtain the equivalent allocated areas, the imputed areas are rescaled by a factor of 2.0 or 2.50, or total physical area to total imputed area, to adjust for the discrepancy of the total area totals. Therefore, the new area totals are 0.32 ha, 0.64 ha, 0.64 ha and 0.40 ha, respectively, which add up to 2.0 ha.

Using the fixed area ratio
In some cases, the number of component crops in the mixture is large, especially when experts consider many of the crops to be important, such that crop area and crop yield estimates are desired for these crops and thus included in the allocation of the total area. Under such case, although the area is recorded at the farm level, the apportionment may be performed at a higher level (for example, the village or district level) using a fixed area ratio determined through eye estimates carried out periodically.

To better illustrate this, the method used in the state of Uttar Pradesh in India is now considered. The important mixtures, such as wheat and barley, are recorded in the village registers as "mixtures"; no attempt is made by the village officers themselves to separate the area among the component crops. Rather, the area is apportioned after aggregating all village-level crop records in the Central Offices. An appropriate formula for each crop is prepared for every district or homogeneous portion of the district, in consultation with the District Officer (CSO, 2007).

Allocating the area when component crops are harvested in different seasons (multiple cropping)
Crops that do not have the same harvesting periods are sometimes planted in the same field. For example, corn, that is normally harvested after seven months is sown at the same time as beans, that are harvested after three months in a parcel. This poses another problem in the estimation of areas apportioned among the component crops. As a remedy, the whole area of mixture is treated as if it were double-cropped; therefore, the area under the component crops are double-counted. Then, the double-counted area is apportioned using any one of the methods outlined above.

Recommended area-apportioning methods

Table 6.1 summarizes the suggested methods for apportioning areas of crop mixture, considering various crop combinations.

TABLE 6.1. SUGGESTED METHOD FOR APPORTIONING AREA BY CROP MIXTURE.

Crop combination	Suggested method of apportioning
Temporary crops only, harvested at the same time Example: wheat and chickpea	Plant density
Temporary crops only, but harvested in different seasons	Area double counted
Associated crops (combination of permanent and temporary crops) Example: mango and sorghum	Entire crop area recorded under permanent crop, area occupied by temporary crop recorded in the season the temporary crop is harvested
Permanent crops Example: coconut and cacao	Area may be apportioned on the basis of number of plants, with adjustment of the plant population of the pure stand of crops constituting the mixture. [is my edit ok?]

Source: Sud *et al*., 2017.

6.2 ESTIMATING YIELD AND PRODUCTION UNDER MIXED CROPPING

Fermont and Benson (2011) give a comprehensive summary of the various strategies that countries use in estimating crop area, production and yield for areas with both pure and mixed crops. The strategies are discussed below. It is assumed that the areas of the fields or parcels on which mixed cropping is practiced, have already been estimated or measured and that the problem is how to estimate the area of a particular crop in the mixture. Moreover, it is assumed that the total areas and average yields mentioned in the following discussions were properly derived. That is, if they were collected using a probability sample survey, then the totals and averages were obtained using the appropriate survey weights.

Strategy 1: Ignore intercropping

Under this strategy, only plots planted to pure stand are recorded for measuring crop area and mixed fields are completely ignored. As a consequence, the true total crop area is underestimated and the average crop yield is overestimated. Thus, for a crop A, total production is estimated as

$$\text{Total production of crop A} = \sum \text{Area crop A}_{\text{pure stand}} \times \text{Average yield crop A}_{\text{pure stand}}$$

Strategy 2: Only record main crop

Instead of completely ignoring intercropped plots, this strategy records only the main or predominant crops, ignoring the minor crops. Estimates of crop area and crop yield are presented as if they are obtained in solely cropped plots, although in reality they are obtained in mixtures of pure and mixed stands. For a given *crop A*, the total area is estimated as the sum of the total pure area of *crop A* and its total area for plots where it is the main crop. While this strategy underestimates the actual area by ignoring the areas where *crop A* is a minor crop, the underrepresentation is significantly lower than occurs with strategy 1. Also, the average yield is determined from a random selection of fields that have *crop A* as a pure stand or as the main crop (for intercropped plots). Thus, the total production of *crop A* for this strategy is estimated as

$$\text{Total production of crop A} = \sum \text{Area crop A}_{\text{pure /main crop}} \times \text{Averag yield crop A}_{\text{pure /main crop}}$$

In the estimation above, the error deriving from not taking into consideration plots where crop A is a minor crop decreases when (1) a sufficiently large number of plots is used to determine the average yield; and (2) crop A is not grown as a minor crop in most of the intercropped plots. Moreover, this method can be easily adapted to give the total area for all crops in either pure stand or mixed stand. However, in reality, some crops generally dominate over the other crops in the mixture. As a consequence, this method may overestimate the actual areas of principal crops and underestimate that of minor crops.

Strategy 3: Use whole plot as a denominator for each crop in the mixture

In this strategy, the entire plot size is used as a denominator for each crop in a mixture for both area and yield estimations. The crop mixture is indicated for each area and yield estimation. The total area for *crop A* is estimated as the sum of the total pure area of crop A and the total area of crop A in all its recorded mixtures, whereas the average yield in this strategy is determined separately for *crop A* as a pure stand and for each of its recorded mixtures.

Capturing the total production for *crop A* would involve inclusion of the area and average yield for *crop A* for all recorded mixtures, which is a cumbersome task to accomplish. To simplify data recording, a criterion may be used to decide whether a particular mixture is included in the estimation, thereby reducing possible recording errors. Kelly *et al*. (1995) suggests including only the two most important mixtures for *crop A* within a region to capture most of the production value. Under this proposition, the total production for *crop A* is then estimated as

$$\begin{aligned}\text{Total production of crop A} = & \sum \text{Area crop A}_{\text{pure}} \times \text{Average yield crop A}_{\text{pure}} \\ & + \sum \text{Area crop A}_{\text{mix\_1}} \times \text{Average yield crop A}_{\text{mix\_1}} \\ & + \sum \text{Area crop A}_{\text{mix\_2}} \times \text{Average yield crop A}_{\text{mix\_2}}\end{aligned}$$

where *mix_1* and *mix_2* represent the most common crop mixture for *crop A*.

When the crop areas used in strategy 3 are summed up at the domain level, the total cultivated area for all crops is likely to be larger than the total agricultural area, because mixed field areas are counted as many times as the number of crops being grown in the sample fields.

Strategy 4: Allocate part of the plot size to each crop in the mixture

In the fourth strategy, the observed area and yield estimations are adjusted to pure stand estimation. This is done by proportionally dividing the plot size between the crops planted in the mixture during both area and yield estimations. Fermont and Benson (2011) note that the area can be apportioned among the component crops in three different ways: (1) eye estimation of the proportion occupied by each crop; (2) examination of the seeding rates or measurements of crop density; or (3) using fixed area ratios for each intercrop combination. As recommended by FAO (1982) and presented in the previous section, the imputed areas of component crops can also be derived using the volume of production or the commercial value of produce.

The total area for crop A is estimated as the sum of the imputed crop areas, and the average adjusted yield is determined from a random selection of plots that have crop A, whether as a pure stand or an intercrop. Therefore, using this strategy, total production for crop A is given as

$$Total\ production\ of\ crop\ A$$
$$= \sum Area\ crop\ A_{adjusted\ to\ pure\ stand}\ x\ Average\ yield\ crop\ A_{adjusted\ to\ pure\ stand}$$

As discussed above, the total imputed areas of component crops may be larger than the area of the mixed field; therefore, the total crop area at a higher level (such as the stratum or domain) may be overestimated. If areas must be aggregated, the corresponding allocated areas should be reported, while the imputed areas should be used to derive the yield. This approach would then eliminate the possibility of double counting and would support the comparability of areas across domains (regions or districts). Because imputed areas are used to compute yields, more suitable yield estimates that reflect that effect of mixed cropping are used.

6.3 INTEGRATING DATA COLLECTION AND ESTIMATION METHODS FOR MIXED CROPS IN NATIONAL AGRICULTURAL SYSTEMS

Countries have different objectives and, therefore, different data requirements for monitoring and formulating policies on crop production. In the case of mixed crops, the NSO may choose to report every possible crop mixture. This approach would be viable only when a manageable number of crop combinations exists across the country. If crop combinations vary across regions, enumerating all possible crop mixtures is not practicable and may overwhelm the users with unnecessary details (Poate and Casley, 1985). Moreover, achieving estimates of yield and area with tolerable precision for each crop mixture may require a larger sample size than the budget available for surveying may allow.

Some countries may require data on particular crop mixtures that are important to their system of production. Also, an area planted to a combination of grain crops may be difficult to apportion, such that countries may treat them as a single crop; an example is "mixed cereals for grain". If this is the recommended practice, then the data collection instrument and survey operations procedure must also be customized accordingly. Enumerators must be alerted to the need to identify the specific crop combinations and must be instructed on recording their data as appropriate. They must also be trained on the procedure for recording other types of crop combinations. It is recommended that where possible, the area of mixed fields should be apportioned into their component crops, to enable meaningful comparisons.

In a general-purpose survey, the farmer is asked to recall or provide an estimate of the production and area of crops grown in the holding. The questionnaire may or may not provide for indicating whether crops are grown in mixed fields or as pure stand. The area in a mixed field may be allocated at the farm level. In some states of India, for example, the village accountant is tasked with apportioning the areas of mixed fields. The allocated areas are then recorded in the questionnaires.

For some countries, major crop combinations are identified and their areas are recorded at the holding level. The allocation of areas of component crops are then done at a higher level, such as the domain or stratum level, using fixed ratios. For unidentified combinations, the entire area is recorded under the major crop and other crops are ignored.

If, instead of farmer recall, objective yield and area measurements are taken for crops in sample holdings, whether under pure stand or mixed crops, then imputed areas for component crops can be derived and used to derive production, which is the product of the yield and the imputed area of crops in the sample holding.

When data collection is done through the village management and only village summary statistics are forwarded to the next level of the hierarchy, crop combinations are recorded at the village level and areas of component crops in mixed fields are allocated accordingly. Yields are derived using the reported production and imputed areas.

Research has shown that when mixed crops are not sufficiently reported, productivity and consequently income will be underestimated. It is therefore prudent for countries to report mixed cropping with at least two levels of detail: (1) the overall land area on which principal crops are grown, together with crop yields and for each crop; and (2) a breakdown of the area into certain important combinations of crops (for example, maize in pure stand, maize with other cereals, maize with beans and pulses, maize with permanent crops or maize with other crops) (Poate and Casley, 1985). This approach eliminates the issues resulting from the impossibility to aggregate crop areas and the improper estimation of crop yields. To derive the total land area for principal crops, the total area of the crop in pure stand and the total allocated area of the crop in mixed fields should be combined. The crop yields, however, should be computed using the total area of the crop in pure stand and the total imputed area of the crop in mixed fields.

Few studies on measuring area and yield for mixed crops could be found and utilized in writing this chapter. The potential for underestimating national production because of the inadequate measurement of crop statistics should be studied further.

7

Methods for measuring crop production or yield under continuously harvested root crops

As outlined in previous chapters, collecting even basic crop production data in low- and middle-income countries and from smallholder farmers presents several challenges.

First, low levels of education often imply that farmers do not keep records of any sort and must rely on recall to report on past events. Furthermore, common cultivation practices such as intercropping, continuous planting and plots of small size and irregular shapes make it particularly difficult to obtain an accurate estimation of smallholder production and productivity. The difficulties inherent in properly estimating crop productivity are further compounded by the length and frequency of harvesting for those crops for which proper quantification of outputs is particularly challenging. This is the case for root crops such as cassava, which are harvested in small quantities over extended periods of time because of better in-ground storability, often spanning even across agricultural seasons. As a result, in this context, the traditional long-term recall methods commonly used in household surveys and the crop cutting methods used in agricultural production surveys may be highly inaccurate and unsuitable.

The importance of root and tuber crop production in low- and middle-income countries cannot be overstated. In 2002, FAO hosted the Expert Consultation on Root Crop Statistics in Zimbabwe, with the ultimate goal of developing methodologies for estimating and forecasting the production of root and tuber crops, in response to the need for methodological progress in this area. The consensus was reached that for continuously harvested crops, production "should be estimated through the periodic measurement of harvested quantities, using local measurement units and conversion factors when standard weights are not applicable"; however, it is necessary to systematically field-test alternative approaches for the robust estimation of area yield for the most important root crops. With respect to land area measurement, the Expert Consultation advocated for the use of GPS units on a more systematic basis.

Sixteen years have passed since the 2002 Expert Consultation, with no major methodological progress to report. While different methods have been proposed and applied in the field, the lack of best practices remains a constraining factor in providing technical advice to countries on correct methodologies. Although crop cutting continues to be considered the gold standard, it is impractical in most large-scale surveys, particularly if they are multipurpose in nature. Furthermore, for most root crops, crop-cutting methods are often not feasible or extremely difficult to implement: for example, although 12 months from planting has been proposed as the best time for estimating cassava production, the timing, time requirements and the high costs render it impractical in most household survey, and even agricultural survey, contexts. On the other end of the spectrum, it is increasingly accepted that using recall methods spanning over several months results in highly inaccurate estimates of continuous crops such as cassava. The deficiencies are rooted in the inability of existing survey methodologies to accommodate continuous harvests that cut across agricultural seasons, with timelines that vary significantly by crop varieties. With an increase in the frequency of unforeseen weather events, governments and donors are strategizing on sustainable coping strategies. In this realm, especially cassava cultivation is now garnering even more attention from policy-makers throughout sub-Saharan Africa due to its vigorous growth, more dependable yield, better performance even under poor soils, relative drought resistance, flexibility to be harvested before maturation, and good storage in field for food security purposes. Additionally, managing cassava fields is not labour-intensive and does not always require fertilizer.

This chapter reviews some of the challenges in measuring production and yield specific to root and tuber crops, details the several methods available and typically used for measuring crop production or crop yield, and provides recommendations to survey practitioners, building on the results of two randomized household survey experiments focused on one of the most important root crops (cassava) and on the recall, diary, and crop-cutting methods. Ultimately, the choice of data collection method depends on available resources, the scale of investigation and the level of precision desired. Therefore the resource requirements, advantages and disadvantages of using each method are presented in this chapter.

7.1 CHALLENGES OF MEASURING ROOT CROP PRODUCTION AND YIELD

While it is relatively simple for farmers to recall harvest information on cash crops, a number of challenges exist in the measurement of continuous crop production (Carletto, Jolliffe and Banerjee, 2015). Cassava is often harvested continuously over extended periods of time and, given its role as a food security measure in the face of food insecurity, the cassava cultivated by a household may not be fully harvested if it is not needed. Additionally, non-standard measurement units may be problematic and misreporting the condition of the crop may further complicate the measurement. In the case of cassava, common units include sacks, heaps or pieces. In a best-case scenario, a farmer may recall the exact number of units harvested over a particular period of time; however, there is enormous variation in terms of the weights of different units, potentially even when reported by the same farmer (for example, a farmer may report two sacks, but one is four-fifths full while another is overflowing).

BOX 7.1. CHALLENGES UNIQUE TO CONTINUOUSLY HARVESTED ROOT CROPS.

Cassava and other continuous crops that are often harvested in small quantities over extended periods of time pose particular challenges for the estimation of crop production and yield, because of the length and frequency of harvesting. The following are some of the main difficulties associated with measuring production in smallholder production systems.

- Farmers do not keep records.
- Recall is widely used in household survey operations; however, it does not always work.
- Crop cutting is more expensive, as well as time- and resource-intensive.
- The use of non-standard measurement units of varying sizes is rampant.
- There are different measurement units along the value chain, and crops appearing in different conditions.
- The development of conversion factors for expressing product-condition-non-standard unit combinations in kg-condition-equivalent terms is still in its infancy.
- Improved approaches to farmer-reported production measurement need validation.

7.2 COMMON METHODS FOR MEASURING ROOT AND TUBER CROP PRODUCTION IN SURVEYS

This section discusses three of the methods for estimating crop production or yield outlined in chapter 5. These are the three most common methods used for measuring root crop production: farmer recall (chapter 5, section 5.1), crop cards (chapter 5, section 5.1.3) and crop cutting (chapter 5, section 5.2.1). The focus is on the advantages and disadvantages of each method as applied to root crops and, in particular, cassava.

FIGURE 7.1. CASSAVA FIELD IN THE UNITED REPUBLIC OF TANZANIA.

7.2.1 Collection of production data through interview: farmer recall and crop cards

In the context of root crops, the prevailing farmer-recall-based approaches to data collection in household and farm surveys are particularly problematic. For cassava, the crop at the centre of this analysis and that is often planted to hedge against the risk of seasonal crop failure or food insecurity during the lean season, harvesting could take place over an extended period and may never be fully complete, depending on the needs of the cultivating household. Therefore, if the survey data collection on crop production is anchored in a specific reference season, the recall-based data on cassava production may be missed or incomplete if the harvest takes place outside the reference season.

If, on the other hand, an extended (commonly 12-month) reference period to capture cassava production is used that does not adhere with the agricultural seasonal calendar, the respondents may not be able to accurately recall the accounts of continuous harvests that take place throughout the year and often in small quantities. The use of non-standard measurement units in farmer-reported crop production further complicates the prospects for quantification in kg-equivalent terms. For instance, common units in which cassava production can be reported include sack, heap and piece, whose weight can vary dramatically across farmers (in terms of time of the year, locality, and cassava variety) as well as for the same farmer (as in the case of a farmer reporting two sacks, of which one may be 80 percent full while the other may be overflowing).

7.2.2 Objective measurement: crop cutting

FAO has recognized crop cutting as the gold standard for yield measurement ever since the 1950s. Besides the cost- and supervision-intensive nature of the exercise, several concerns have been raised regarding the accuracy of the method. Even if one places only one random crop cutting subplot within the sampled plot, the resulting yield estimate may carry a sampling error if the yields exhibit within-plot heterogeneity. Several additional potential sources of error have been put forth, including (1) more thorough harvesting of crop-cut subplots vis-à-vis typical farmer harvesting practices; (2) possible rounding of crop-cut production estimates obtained through scales; (3) using faulty or inappropriate scales; (4) omitting to net out the weight of the measurement container from the measured production; (5) including plants that fall outside of the subplot; and (6) non-random placement of crop-cut subplots (Fermont and Benson, 2011).

7.2.3 Comparison of existing methods

Few studies have been conducted to compare recall versus diary approaches in continuous root crop agricultural data collection.[1] In an approach assumed to reduce the potential biases resulting from recall-based production estimates, Deininger *et al*. (Deininger, Carletto, Savastano and Muwonge, 2012) is the first and only study comparing recall and diary estimates for cassava, captured as part of a larger crop diary operation held in parallel with the Uganda National Panel Survey (UNPS). The magnitude of discrepancies between the two methods was especially high for cassava, in terms of both the households reporting cultivation of cassava at all, and of the value of production. Recall underestimated the percentage of households involved in cassava cultivation by 24 percent and the value of production by approximately 14 percent. The study acknowledged that even with these already large discrepancies, it was possible that even the diary arm underestimated cassava production: the duration of the diary operation was six months, while cassava can be harvested for a much longer period of time.

1 The body of work investigating the accuracy of recall data and diary methods is a much broader area of study beyond agricultural production data. Beegle *et al*. use a field experiment in the United Republic of Tanzania to compare eight different methods for collecting consumption data, finding that recall modules measure lower consumption than a personal diary, with larger gaps among poorer households. (Beegle, De Weerdt, Friedman and Gibson, 2012).

To address the gaps in the literature on continuous crop production and yield data collection methods and integration in household surveys, the World Bank Living Standards Measurement Study (LSMS) has prioritized the most important continuous crop, cassava, in its research agenda. The main recommendations in this chapter are based on two methodological validation studies conducted in the United Republic of Tanzania (Zanzibar) and Malawi. To document the relative accuracy of existing survey methods, field experiments were implemented in these two countries over 12-month periods, randomly assigning the sampled households in key cassava-producing areas to one of four approaches to cassava production measurement: daily diary-keeping, with semi-weekly in-person supervision visits (D1); daily diary-keeping, with semi-weekly supervisory telephone calls (D2); a two-visit, six-month recall-based data collection effort, with visits six months apart (R1); and a single visit, 12-month recall-based data collection effort (R2).

BOX 7.2. DESCRIPTION OF LSMS-ISA METHODOLOGICAL EXPERIMENTS IN UNITED REPUBLIC OF TANZANIA AND MALAWI.

A. Data for Zanzibar, United Republic of Tanzania come from the Measuring Cassava Productivity (MCP) study. The MCP focused on testing several methods for measuring cassava production, complemented by the land area measurement of cassava plots using three different methods. The fieldwork was done from June 2013 through May 2014, with area measurement being conducted between August 2013 and January 2014. The study was carried out in two districts, one on Unguja Island and one on Pemba Island. The enumerators were local agricultural extension officers. The sample consisted of 1 247 households, with 1 932 cassava plots measured for land area. Partners in the study included the Ministry of Agriculture and Natural Resources of Zanzibar, the Office of the Chief Government Statistician of Zanzibar, and the World Bank. The handheld GPS unit used in the study was a Garmin eTrex 30.

B. The data set for Malawi comes from an experiment formally known as CVIP: Methodological Experiment on Measuring Cassava Production, Productivity, and Variety Identification. This experiment was implemented between July 2015 and August 2016 in 45 EAs across the top five cassava-producing districts in the Central and Southern Regions of Malawi. Cutting across different districts allows the CVIP to capture multiple cassava production systems and a range of cassava varieties.

C. In each of the studies, cassava-producing households were randomly assigned to four survey treatments that differed in terms of their approach to data collection on production. The approaches taken were: (1) daily diary-keeping (D1) at the household level for 12 months with in-person supervision visits held twice a week; (2) daily diary-keeping (D2) at the household level for 12 months with supervisory mobile phone calls twice per week; (3) recall-based data collection (R1) at the plot level for a six-month reference period, administered twice during the fieldwork with visits after 6 months and again after 12 months; and (4) recall-based data collection (R2) at the plot level for a 12-month reference period, administered only in the last month of fieldwork (the prevailing practice in household and farm surveys collecting information on cassava production). For every survey household, one cassava plot was selected at random for crop cutting, and a 5 m x 5 m subplot was set up in line with international best practices. Diary households were given scales to measure their harvests in standard units.

The full empirical framework for estimating the relative survey treatment effects and resulting analysis is outlined clearly in *The Root of the Measure: Cassava Productivity in Zanzibar* (Banerjee, Carletto and Mzee, ongoing work) and *Extended-Harvest Crop Production and Productivity Measurement in Surveys: Results from a Randomized Field Experiment on Cassava in Malawi*" (Kilic *et al.*,2018]) Intra-EA randomization of the systematically sampled households across treatment arms in the two studies allowed the authors to estimate the causal effects associated with each treatment arm on production and yield.

The authors of the two studies find that if true cassava production is assumed to be underestimated even by diary-keeping because of imperfect compliance, D1, despite being traditionally considered the gold standard, underestimates annual production, while the D2-based estimates may in fact be closer to the true production value despite the commonly held beliefs around the importance of the personal supervision of respondent-kept production diaries. Both recall variants underestimated, relative to D2, annual production by a significant margin, although R1 did so to a lesser extent. Furthermore, given the lack of variation in the total household area cultivated by cassava across the survey treatment arms, the yield comparisons among the diary and recall methods mirror the results emerging from production.

In Malawi, the average household-level crop-cutting-based annual cassava yield estimate stands at 8 958 kg/ha across all sample households, while the comparable D1, D2, R1 and R2 estimates are 5 208, 6 218, 5 798, and 4 671, respectively. The implication is that the diary and recall survey treatments, on average, underestimate the yield compared to crop cutting. The extent of underestimation, as a percentage of the crop-cutting-based yield, is of 40 percent, 25 percent, 33 percent and 47 percent for D1, D2, R1 and R2, respectively. Regardless, the average CVIP crop-cutting-based yield estimate is significantly lower than the annual FAOSTAT Malawi estimates of 24 800 kg/ha in 2014, as well as the MoAIWD APES-based 2015–2016 national cassava yield estimate of 17 564 kg/ha. More specifically, the FAOSTAT estimate is almost three times the size of the crop-cutting-based yield estimate, and four times the size of the D2-based counterpart.

7.3 KEY CONSIDERATIONS FOR SURVEY PRACTITIONERS

In household and farm surveys across low- and middle-income countries, the prevailing approach to data collection on the production of root crops is to field a single visit and collect information that is either specific to an agricultural season or with a 12-month recall. In turn, there are concerns that the existing survey methods may not elicit reliable information on crop harvest cutting across seasons, taking place over extended periods, and varying by variety.

BOX 7.3. COST COMPARISONS OF THE LSMS-ISA METHODOLOGICAL EXPERIMENT CONDUCTED IN MALAWI.

The results from the two methodological validation studies provide evidence based on which survey practitioners should move away from R2, given the significant underestimation of production and yield estimates, and assert that the D2 households provided the most accurate information on annual cassava production. The latter is shown to be ensured through sustained participation in diary-keeping, which could have been connected to the relatively more valuable set of in-kind incentives these households received to be part of the experiment – although this hypothesis is not testable. However, in Malawi, while D2 is significantly cheaper compared to the traditional diary operation, it is still more resource- and supervision-intensive compared to the recall variants. The D2 unit cost per household is USD 330, compared to USD 469 for D1, USD 186 for R2, and USD 157 for R1.

Thus, even if implementing crop diaries with telephone supervision is too costly for the survey implementing agencies, it would be more feasible if it were packaged within a broader effort to collect more frequent data through mobile phone calls on a wider range of topics, including (but not limited to) extended-harvest crop production, following baseline face-to-face interviews. In the case of Malawi's NSO, an effort such as the existing Listening to Malawi (L2M) mobile phone survey, as part of the regional Listening to Africa initiative (Dabalen *et al.*, 2016), could provide such a platform. In the worst-case scenario, the analysis indicates that R1 is the second-best alternative to D2 and a clear improvement over R2. R1 performs as well as the traditional diary operation (even if it underestimates annual cassava production with respect to D2), and entails an additional cost per household that is USD 30 lower compared to R2.

TABLE 7.1. COMPARISON OF EXISTING METHODS FOR MEASURING CROP PRODUCTION OR YIELD.

Acronym	Method	Cost-effectiveness	Precision in estimation, errors and biases
D1	Daily diary-keeping at the household level for 12 months, with semi-weekly in-person supervision visits	Cost- and labour- intensive	Fairly accurate estimation; however, needs adequate supervision due to underestimation with respect to diary-phone.
D2	Daily diary-keeping at the household level for 12 months, with semi-weekly supervisory mobile phone calls	Significantly cheaper compared to the traditional diary operation, it is still more resource- and supervision-intensive compared to the recall variants	Provided the most accurate information on annual cassava production and yield estimates.
R1	Two-visit recall-based data collection at plot level for a six-month reference period, carried out in two visits six months apart over a 12-month period	Cheaper and quicker method that saves time and money in comparison to diary operations	Second-best alternative to diary-phone and a clear improvement over 12-month recall. Performs as well as the traditional diary operation (even if it underestimates annual cassava production with respect to diary-phone).
R2	Recall-based data collection at plot level for a 12-month reference period, administered in a single visit	Cheap and quick method that saves time and money in comparison to diary operations.	Significant underestimation of production and yield estimates.
CC	Crop cutting	Cost- and labour- intensive.	Significant overestimation.

To expand on how diary-phone operations or six-month recall could be operationalized in the context of existing household surveys, three survey types are considered:

1. **Single-visit survey**, which could inquire about the last 12 months or a given season when it comes to extended-harvest crop production;
2. **Two-visit survey**, which would entail a post-planting and a post-harvest visit around a main agricultural season, and could inquire about the last 12 months or the reference agricultural season when it comes to extended-harvest crop production; and
3. **Two-visit survey**, which would entail two post-harvest visits for two agricultural seasons, and could inquire about the last 12 months in one of the visits or about the reference season in each visit when it comes to extended-harvest crop production.

Single-visit surveys eliciting information on agricultural production typically focus on cassava harvested in the 12-month period preceding the interview, or include cassava under rainy-season crops, thus phrasing questions on continuous crop production as if crops were harvested at the end of the main agricultural season.

For a survey of this nature to adhere to the recommendations made in this body of work, many adjustments to the survey operation protocol would have to be made. If the survey was well-funded and survey implementers may potentially afford a mobile phone effort to run in parallel with the main fieldwork, Dabalen *et al*. (2016) note several considerations that would still need to be considered including the sampling design, fieldwork duration, and implementation of a baseline survey beyond the costs of establishing and maintaining a call centre, as well as the incentives provided to survey respondents.

While the overall sample size for this type of survey is likely to have already been set, one could consider following the D2 approach in a subsample of households, depending on the funding available and the representativeness demands from the subsample. After finalizing the sample size, the three major costs will stem from the baseline survey: the hardware required, and the costs of setting up and maintaining a call center. Given the objective of keeping the diaries in the field for 12 months, ideally, the baseline survey administered to the selected D2 households would have to be completed within a month of the start of the main survey fieldwork.

Dabalen *et al*. (2016) further highlights several potential cost-saving mechanisms, including the existence of another visit to households that could serve as the baseline; providing mobile phones only to households without a mobile phone; and the availability of equipment at the NSO headquarters to establish and run the call centre. An option for a baseline for the D2 households could be part of the listing exercise. If the survey is implemented using CAPI, household selections can be done while listing teams are still in the EA and cassava-producing households could be identified and interviewed further. Depending on the structure an institution has established for running its listing exercises, this is likely to require a more intensive training for listing and additional days in cassava-producing EAs. However, the money saved from taking advantage of enumerators that are already within these EAs has the potential to outweigh the aforementioned costs. This consideration also relies on the assumption that the listing exercise is completed prior to the main fieldwork, but also not excessively in advance of the start date as the call centre should begin its operations at the same time.

Two visit surveys, whether collecting data on one main agricultural season through post-planting and post-harvest visits or collecting data on two agricultural seasons, may still vary greatly in their approach in terms of timing. The approach shown above should be taken to potentially implement the diary-phone approach to data collection, knowing that because enumerators will already be visiting households twice, there may be at least minimal cost-savings. The implementation of R1, however, may be treated differently in a two-visit survey.

To present a clearer picture of the types of surveys envisioned for these recommendations, the Uganda National Panel Survey (UNPS), supported by the World Bank LSMS-ISA initiative, provides an ideal set-up for implementing the R1 operation, given the six-month spread between the two visits, each of which solicits information on a six-month agricultural season. In contrast, the Malawi Integrated Household Panel survey series, despite also having a two-visit post-planting and post-harvest set-up, is not immediately suitable for the R1 approach to data collection because the visits occur, on average, only three months apart from one another. Where the timing of the visits is not in line with the six-month recall set-up, survey practitioners will need to consider including an additional visit.

8

Estimating area and yield of horticultural crops

Vegetable cultivation is one of the most important components of horticulture, which is currently among the fastest growing sectors in most developing countries. Compared to traditional field crops, horticulture presents several advantages that make it attractive to both commercial and subsistence farmers.

First, horticultural crops, especially the vegetable ones, serve a dual function of cash and food crops, contributing significantly to household income and food security. In this sector, smallholdings contribute substantially to the market, horticulture production being an activity that can also be performed in small portions of land.

In addition, enormous opportunities to augment the exports of horticultural commodities and products are emerging, thanks to the strengthening and better organization of existing and new markets.

Finally, the horticultural sector presents great potential in terms of employment, as it requires a series of direct and indirect activities related to land preparation, nurseries, cultivation, harvesting, post-harvesting, trading, storing, processing, transportation, marketing and distribution of horticulture commodities and products. The expansion of this sector could significantly contribute to reducing unemployment, both in rural and urban areas.

Despite the social and economic importance of the sector, information on area planted and harvested, quantity produced and market prices remains scarce, preventing the design of sound development programmes. In most countries, the quantity produced each year is largely unknown, also because of objective difficulties in properly estimating horticultural activities. Indeed, many technical problems need to be addressed when producing statistics on vegetable crops.

First, the existence of a great variety of vegetables, with differences in phenology and cultivation techniques, makes it difficult to identify a unique methodology to be applied in all circumstances. Furthermore, the definition of the population of interest may also not be straightforward, given the diversity of the holding types currently producing vegetables. In addition, the approaches available to harvesting vegetable crops may differ substantially, each entailing different challenges in terms of statistical measurements. Finally, vegetable production is normally devoted to two main types of utilization: fresh consumption and processing. Consequently, the yield of these two types of vegetable production may differ substantially.

FIGURE 8.1. VIEW OF VEGETABLES GROWING ON A FAMILY FARM IN INDONESIA.

The present chapter examines the methodological options available to estimate vegetable crop area, production and yield with an acceptable level of accuracy. A key point stressed throughout this chapter is that, before implementing any data collection activity, it is paramount to carefully assess the existing data sources (for example, administrative data). As for any data collection exercise, the objectives need to be stated, clarifying the list of data items, the vegetables crops to be included, the scope and the coverage of the survey. Furthermore, the concept of "agricultural holding" must be unambiguously defined, setting proper thresholds or referring to existing legislation, to clearly separate the household and non-household sectors. Finally, any methodology developed should be field-tested before implementation.

BOX 8.1. RELEVANT CONCEPTS AND DEFINITIONS FOR VEGETABLE CROP PRODUCTION STATISTICS.

In the following paragraphs, some concepts related to vegetable crop production statistics will be repeatedly used. These are:

Vegetable crop holding: irrespective of the destination of the vegetable production, any agricultural holding producing vegetables can be defined as a "vegetable crops holding". When applying this concept to data collection, the national definition of holding must be applied.

Types of vegetable crop holding: there are many types of vegetable holdings, according to the size of their vegetable production, the purposes of production and the cultivation tecÚiques adopted. Furthermore, the generic categorization proposed here must be carefully adapted to the specific country context.

- **Small-scale holdings**: holdings with a low production of vegetables. Their size is close to that of a kitchen garden and, generally, they produce many vegetables in a small area without specialization. There are two main categories of small-scale holdings:
  - **Small-scale holdings producing for self-consumption**: holdings producing mainly (80 percent or more of their production) for household consumption.
  - **Small-scale holdings for sale**: holdings that produce mainly for sale, even if a part of their production is consumed by the household.
- **Market gardening holdings**: one of the activities of these types of holding is to produce vegetables for sales in fresh, directly to the consumer or to retail stores or wholesalers.
- **Production of vegetables for industry**: holdings that produce vegetables for canning or deep-freezing industries. Very often, production and marketing are performed under a contract between the holding and the industry.

Horticultural crop holding: the productions of these holdings are bulb cultures (bulbs, onions, tubers, rhizomes and claws), floral nurseries, cut flowers and foliage, plants in flower pots and green plants, plants with rootstocks or clumps and perennials. This type of production is normally carried out for sale and holdings practicing this activity are usually specialized ones.

Bed: the smallest piece of area where vegetables are planted, with paths all around. Paths are necessary to water and weed without having to tread on the bed. When the soil is poor or does not drain well, beds can be built and raised. The sum of bed areas constitutes the area in the relation between production, yield and area.

Developed area: the total area used to produce vegetables. This is the area that needs to be considered in the computation of yield. To clarify this concept, let us consider the example of a holding producing radishes each month over six months on a bed of 200 m^2. The Agricultural Area Utilized (AAU) to produce radishes is 200 m^2; however, the developed area to take into account in the calculation of yield is 200 m^2 x 6 = 1 200 m^2. This developed area can be the sum of several beds, each one used once or more.

Crop rotation: a method of farming in which a number of different plants are grown one after the other on a field so that the soil remains healthy and fertile (that is, able to produce crops).

⏩ ⏩ ⏩

Crops with simultaneous harvest: the entire crop production is harvested at the same time and a new plantation must be done to obtain a new production. This does not mean that, during a specific agricultural year, only one harvest can take place. On the contrary, this kind of crops can be harvested multiple times; however, after each harvest, the crop must be replanted. An example of this type of crop is lettuce.

Continuous harvesting: the crop grows and matures progressively (tomatoes, for example) and is gathered progressively during the entire period of growing. Multiple crop cycles during an agricultural year are also possible.

Staggered crop: staggered planting occurs when the same vegetable is grown, but the seeds are planted on different dates throughout the season to enable a longer period of fresh vegetables. This tecÚique is mainly used in small-scale holdings.

Growth season: the growing season in a particular country or area is the period, during each year, when the weather and temperature are right for plants and crops to grow.

Growth cycle: for each plant, the period of growth between sowing or planting and harvest. There are several vegetative phases during the growing cycle; for example, germination and flowering. Generally, the harvest is carried out during the developing phase with a maximum yield. Some crops could complete multiple growing cycles during one agricultural year.

Two main challenges arise when surveying vegetables producers:

1. Measuring the developed area could be difficult. The size of the area planted could be unknown to the farmer. Furthermore, the identification of beds could be problematic, with many crop varieties often grown simultaneously on the same bed. In such scenarios, specific rules and procedures must be put in place. For example, if three crops are growing on a same bed, one-third of the bed area might be allocated to each crop, or the surveyor may be in charge of allocating the area, with the collaboration of the farmer.
2. The absence of good recordings of farming operations and relevant dates, quantities produced, final destinations of production, and agricultural practice such as use of fertilizers, pesticides and irrigation, makes it difficult to obtain an accurate picture of the vegetable production sector. Lack of recordings is generally a source of underestimation of productions, particularly when there is more than one harvest on a bed.

To gather good-quality and timely data, respondents should be followed during the entire reference period of the survey and encouraged to keep records of their activities, reducing the rate of missing data. A possibility could be followig respondents with phone calls, associated with one or two in-person visits during the reference period.

In this chapter, ornamental plants are considered as a special case of horticultural crops. Indeed, estimating ornamental plants production poses less challenges compared to vegetables crops. This is mainly because normally, holdings producing ornamental crops are highly specialized and produce mainly for the market instead of own consumption. As most of the product is sold to wholesalers, farmers generally keep records of their activities and data for this subsector are often more accurate. The main problem in this context is the definition of an adequate statistical nomenclature that may have to be more detailed than the one used by the holdings interviewed.

8.1 TYPES OF VEGETABLE AND ORNAMENTAL CROPS

Vegetable crops may be divided into three main categories: leafy, fruit and root vegetables. In this section, a fourth category for ornamental plants was added.

Leafy vegetables

Leafy vegetables, also called potherbs, greens, vegetable greens, leafy greens, or salad greens, are plant leaves eaten as vegetables. Although they come from a very wide variety of plants, most share a great deal with other leafy vegetables in terms of nutrition and cooking methods. Almost 1 000 species of plants with edible leaves are known. Leafy vegetables most often come from short-lived herbaceous plants, such as lettuce and spinach.

This category of vegetable crops can be further subdivided in two classes:

- Leafy vegetables with staggered harvest: these are crops for which products do not mature at the same time and for which multiple harvests are required and spread over a certain period of time.
- Leafy vegetables with simultaneous harvest: these crops are completely harvested upon maturity and the plant is destroyed thereafter.

FIGURE 8.2. FIELD OF DIFFERENT LETTUCE VARIETIES IN FRANCE.

Fruit vegetables
Fruit vegetables are edible plants that are botanically considered as fruits but are consumed as vegetables. The category includes seaweed, sweet corn, pulses and mushrooms; and they are generally consumed raw or cooked (tomato, melon, chayote, cucumber, eggplant, okra, papaya, chilli pepper, pumpkin, etc.). The distinction between staggered harvest and simultaneous harvest crops holds in the case of fruit vegetables too.

Root vegetables
This category contains any fleshy edible underground roots or tubers used as a vegetable, such as carrot, potato, celery, onion, parsnip, radish and beetroot. Normally, root vegetable crops (which are different from root crops, such as cassava, discussed in chapter 7) are simultaneous harvest crops that are completely harvested upon maturity and destroyed thereafter.

FIGURE 8.3. POTATO FIELD IN PERU.

Ornamental plants
Ornamental plants are non-edible plants used as ornaments, such as bulbs (bulbs, onions, tubers, rhizomes and claws), floral nurseries, cut flowers and foliage, plants in flowerpots and green plants, plants with rootstocks or clumps and perennials.

Vegetables and ornamental plants can be cultivated using techniques with a different degree of yield orientation, that is, reliance upon different combinations and quantities of land cover, irrigation, fertilizers and plant protection products (PPPs).

The great variety of vegetable cultivation techniques must be taken into account in yield calculation. Yields obtained by professional holdings in vegetable production are not comparable to those obtained by households for their own consumption.

Horticultural crops cultivated in open-air fields. The great majority of vegetable and ornamental plant production are carried out on soil without any cover, although it is normally irrigated. Vegetables in open fields (for example, for industry or market gardening) or ornamental plants can be carried out under low shelters. This kind of cover can be easily transferred from field to field.

Horticultural crops grown in greenhouses. A part of vegetables for consumption in fresh and of ornamental plants grows under greenhouses, which may be heated or not heated but are always irrigated. A greenhouse is a fixed building where temperature and aeration are managed following plant needs. A subcategory of crops grown in greenhouses are hydroponic cultivations. Hydroponic plants are grown in greenhouses without any soil and are constantly fed by a nutrient solution that is pumped from a tank (reservoir) and is most often continually recycled throughout the system. Vegetables such as lettuce, tomatoes, peppers, beans, and cucumbers are all suitable hydroponic crops. Hydroponic crops mostly include food crops rather than ornamental crops such as flowers; however, many keen hydroponic growers are starting to grow flowers in addition to edible crops.

Additional dimensions related to the adopted cultivation techniques that should be surveyed are summarized in box 8.2 below.

BOX 8.2. ADDITIONAL ITEMS TO BE SURVEYED TO ESTIMATE THE AREA AND YIELD OF HORTICULTURAL CROPS.

When conducting a survey on vegetable crops, collecting information on the developed area and the quantity harvested may not be sufficient to properly estimate the final yield.

The survey questionnaire should include questions on the following items:

- size of land with greenhouses or hydroponic systems;
- size of land with open-air fields;
- size of irrigated land, drained land in open-air field;
- land under certified organic farming and land in conversion to organic;
- type and quantity of fertilizers (mineral, organo-mineral, compost, mulch, organic, animal effluents, etc.); and
- type and quantity of plant protection products (PPPs) (insecticide, herbicide, fungicide, rodenticide, etc.).

8.2 STEPS TO BE IMPLEMENTED BEFORE DATA COLLECTION

As vegetables are a highly complex category of crops to follow, the objective, scope and coverage of the survey must be clearly defined. The following are some of the main aspects that need careful consideration when developing the data collection methodology for vegetable crops.

List of vegetables to be included

Surveying all vegetables cultivated in a country is often not feasible (except when an agricultural census is conducted). Therefore, the list of vegetables of interest to be included in the survey should be established, considering the need of targeted data users. In cases where the cultivation of a few specific crops are of particular relevance for the country economy, data collection should target at least these vegetables. Other minor horticultural crops could be included if the available budget and frame allow.

Definition of the population of interest

As in any other survey, the population of interest must be clearly defined. In most cases, when identifying agricultural holdings, the main decision to be taken is whether to include households cultivating only kitchen gardens. This decision must consider other surveys, such as household surveys, that are carried out in the country. For example, if a household survey that includes all household typologies and collects information on kitchen gardens already exists, it is not necessary to include kitchen gardens in the vegetable crop survey.

8.3. METHODS FOR DATA COLLECTION

If yield and area are the variables of interest, the targeted agricultural holdings on which area and quantities produced are to be measured must be clearly defined. On the other hand, when the objective is to gather data only on quantity produced and farm gate prices, information can be collected at the market level or from other commercial networks. Information from wholesale and retail markets can also be gathered from administrative sources. When data are needed on a monthly or quarterly basis, it is sufficient to follow the quantity produced and marketed and prices on the market. Doing so, the enumerators will record data on quantities bought and sales prices. This operation should be performed daily and the row data should then be aggregated over the reference period.

Estimating the production and yield of vegetable crops from kitchen gardens or from market-oriented holdings entails different strategies, and the main issues arising could be significantly different. For this reason, this section will separately illustrate these two scenarios.

8.3.1 Data collection methods for vegetable crops grown in kitchen gardens

The main data sources to be used in this case are censuses, sample surveys such as agricultural production surveys and household surveys, or village surveys.

Censuses

As mentioned in chapter 2, population and agricultural censuses could be potential sources of data on vegetable crops:

- **Population censuses** often include questions on the existence of a kitchen garden and its size. Generally, the amount of information provided is not detailed; however, it allows for estimating the total area devoted to kitchen gardens and provides a sampling frame for the identification of this category of producers. Data on quantity produced, instead, are rarely collected through population censuses.
- **Agricultural censuses** often allow for collecting information on the size of land devoted to kitchen gardens, along with information on the size of the other components of the Agricultural Area Utilized (AAU). Furthermore, agricultural censuses can be used to build the sampling frame for other agricultural surveys. However, agricultural censuses normally cover only households in the agricultural sector, while population censuses include all households irrespective of the presence of any agricultural activity. Furthermore, in this case too, data on production are rarely collected (see chapter 2).

Sample surveys

Sample surveys on households growing vegetable crops may rely on list frames developed through one of the two census categories mentioned above. The principal information required to collect with a survey are:

1. the list of vegetables cultivated in the holding; and
2. the size of area planted and harvested, the quantity produced and, when possible, the share sold of each crop.

Furthermore, information on the agricultural practices adopted by the household (use of fertilizers, irrigation, PPPs, etc.) can also be added to the survey questionnaire.

"Villages" surveys

This type of survey is often a means to reduce survey costs. Usually, these surveys are based on a sample selected in two stages, where the first level of selection is normally the "village". The stratification of "villages" can be done based on the region, type of "village" and population number. The second level is the household that can be stratified using the area of the kitchen garden. It is essential to anticipate the organization of this type of survey to include questions on the existence of a kitchen garden and its area in population or agricultural censuses.

8.3.2 Data collection methods for market-oriented vegetables and ornamental plants production

There are three main ways to gather information about vegetables and ornamental plants:

1. integrating a specific module on vegetables into a broader agricultural survey or census;
2. developing specific surveys on vegetable and ornamental crops; and
3. relying on other data sources, such as administrative data, market surveys and remote sensing.

Reporting the exact size of beds, total quantity produced and the share of horticultural production devoted to own consumption could be difficult for respondents. Therefore, the survey or census questionnaire must be carefully tested and adapted to the specific survey situation.

Agricultural censuses

Agricultural censuses may be a means to collect data on quantity of vegetable and ornamental crops produced, area planted and area harvested. Usually, visiting a holding more than once during a census is not feasible, and all the information is collected during one visit by interviewing the holder of the agricultural holding. The list of all holders interviewed could then serve as a sampling frame for an annual agricultural survey.

Module in an annual agricultural survey or ad hoc survey

The selected sample should include all types of holdings producing vegetables or ornamental plants. When an annual survey is implemented, some strategies to improve the quality of answers received may be put in place, particularly for market gardening holdings. As it is often difficult for respondents to provide information on area and production, an option to improve data quality consists in using journals or diaries to keep record of their activities (an example of diary is presented in annex 10). During the first visit, the surveyor and the respondent must determine the list of beds and their size. In doing so, a properly trained enumerator can rely on the use of GPS for the largest beds, while simple measurement could be sufficient for the small ones. This procedure could help to obtain quality area data when the respondent is unaware of the exact size of land. A map can be drawn to help the respondent memorize the different beds.

After these preliminary operations, bed characteristics such as cover and irrigation, the types of crop planted, and other information can be recorded. When surveys entail the compilation of a journal, the enumerator must explain to the respondent how to regularly record his or her activities. Variables normally recorded in a diary are:

- date of sowing or planting;
- crops planted;
- harvesting dates;
- quantity harvested;
- harvest destination; and
- selling price or total revenues obtained.

Where the number of beds is excessively high, a random selection can be done to reduce the respondent burden. The surveyors must be well prepared to perform this task as it usually cannot be prepared before the beginning of the survey. Additional questions for the diary could cover the use of inputs, pesticides, storage facilities and other equipment, and commercial networks.

The respondent must be regularly contacted, by telephone for example, to remind him or her to fill in the diary and to answer the questions. If possible, an intermediate visit could be organized.

After accessing data recorded in the diary, all aggregated information should be validated. A visit to the holding should be organized at the end of the survey period, during which the enumerator should validate information recorded for each vegetable. This operation entails visiting each selected bed and assessing data for each vegetable grown in it.

When vegetable crops are grown in open-air fields, it is possible to collect all data during the same visit when information on other crops (such as cereals) are collected.

With regard to ornamental plants, it is generally possible to ask questions on the entire survey period.

Pilot tests conducted before survey implementation should be oriented to establish a nomenclature with the right level of detail, taking into account respondents' ability to provide information on the quantities produced.

Use of administrative data

In some countries, statistics on vegetable crops may be also compiled using administrative data sources. Indeed, in most countries, administrative data on input use (such as seeds and pesticides), market supply, and import and export of vegetables is often available.

The advantage of relying on administrative data sources is that they are usually already available for free-of-charge use. However, it is necessary to evaluate administrative data quality in terms of coverage, exhaustiveness and accuracy.

Use of other sources

When implementing a market survey, survey enumerators visit markets where vegetables are sold to record quantities sold and market prices. The most viable strategy consists in appointing a correspondent from each market who is capable of transferring the collected information to the statistical office in charge of the survey.

Another potential source of data consists in the use of remote sensing. This approach presents serious difficulties when the interest is for vegetable and ornamental plant production, especially because it cannot be used for vegetable crops grown in greenhouses. Furthermore, also for small holdings cultivating vegetable crops in open-air fields, the main challenges to face in implementing remote sensing are the following:

- The size of beds may be small compared to the definition of the smallest polygon that can be observed.
- Beds could be covered by low shelters or greenhouses where the estimation of the covered crops is impossible.
- It could be difficult to separate vegetable crops from other crops following the response it gives and the level of maturation.

TABLE 8.1. COMPARISON OF ALTERNATIVE DATA COLLECTION METHODS FOR VEGETABLE CROPS.

Survey method	Pros	Cons	Cost-effectiveness
Agricultural census	Completeness. It also enables creation of a sampling frame of holdings or households growing crops.	Provides information only on area cultivated per vegetable. Implemented infrequently.	Collecting information on vegetables has a marginal impact on the overall census budget
Module in an annual agricultural survey	Annual results. Main aspects of vegetable production can be studied. Microdata from an annual survey can be used for more complex analyses and research.	Increase interview duration. Necessary to adapt the sample to ensure representativeness.	Annual survey budget slightly increased.
Ad hoc survey	More frequent results than agricultural censuses. Specifically targeted to vegetable crops area and production.	Organization of a supplementary survey. Greater burden on NSOs.	Necessary to find supplementary budget for a new survey.
Administrative data source	Free-of-charge and already exists.	Possible problems of data quality and definitions. Possible changes in administrative objectives and risk of lost information or changes in definitions.	Low budget
Market survey	Possible high frequency	Only collects information on quantity sold and prices.	Low budget
Remote sensing	No field survey	Difficulties in photo-interpretation. Greenhouses cannot be captured.	Depends on tecÚologies and skills available in the country.

8.4 ESTIMATION OF DEVELOPED AREA AND PRODUCTION

The correct method for the estimation of the developed area and production of vegetable crops depends on the specific crop category considered (leafy, fruit, or root vegetables with single or multiple harvests).

In case of a two-level sample, if the PSU is the holding and the SSU is the bed, the developed area for each vegetable must be extrapolated at holding level, that is, for all beds cultivated by the holding. This operation is achieved using a coefficient that is defined as the total developed area of selected beds cultivated with the considered vegetable, divided by the total developed area of all the beds selected. In formulas, the coefficient for vegetable V will be:

$$\frac{\sum_{j=1}^{n} AB(j)^V \times NAB(j)^V}{\sum_{j=1}^{m} AB(j) \times NAB(j)}$$

where

$AB(j)^V$ = net area of bed j harvested at least once for vegetable V during the reference period

$NAB(j)^V$= number of harvests of vegetable V during the reference period on bed j

n = total number of beds where vegetable V has been harvested at least once during the reference period

m = number of sampled beds

Production of leafy vegetables
The production of leafy vegetables can be declared in terms of weight or of number of plants. In this last case, the weight must be calculated applying the average weight of plants provided by the holding. If a sample of beds has been selected, the total production of each vegetable at the holding level must be extrapolated using a similar coefficient as the one developed for the area. In case of multiple growth cycles during the reference period, the entire production for each harvest must be taken into account.

Production of fruit or root vegetables
Production of fruit vegetables is generally declared in terms of weight. It may be necessary to convert local units into standard units of measure. In this case too, if a sample of beds has been selected, the total production of each vegetable must be estimated in the same way as the area. In case of multiple growth cycles during the reference period, the entire production for each harvest must be taken into account.

8.5 EXAMPLES FROM SELECTED COUNTRIES

8.5.1 Horticultural data in the United Kingdom

The June Survey of Agricultural and Horticultural (JSAH) Activity has a long-standing history in England. First run in 1866, it has been carried out as a full census every year until 1995, when it became a survey with a sample approximately covering the 80 percent of the farming population. Today, the survey samples between 30 000 and 70 000 holdings each year. Full censuses are carried out every ten years, with the next one expected for 2020.

The JSAH is conducted with the objective of collecting detailed information on arable and horticultural cropping activities, land usage, livestock populations and agricultural labour force figures.

The sample survey is run every year on 1 June across England. It is mainly postal, although farmers are also invited to complete it online through the Whole Farm Approach (WFA). The survey adopts a stratified random sampling in which holdings are divided into strata based on their "theoretical labour requirement", with a higher sampling rate being used in the larger strata. The Standard Labour Requirements (SLRs) are calculated using coefficients across all activities of a holding to give a standardized measure of labour activity and an indication of the number of full-time workers required to run a holding. Simply stratifying on labour requirements would lead to unacceptably low precision for certain important crops, especially horticultural crops. Therefore, strata with higher sampling rates for horticultural holdings have been separated.[1]

In recent years, the Department for Environment, Food and Rural Affairs (DEFRA), the government agency in charge of the survey, fixed a predefined threshold to exclude small farms from the survey. Therefore, the survey population now only covers the larger and more market-oriented agricultural holdings. On the other hand, national estimates of the activities of holdings falling below the established threshold are computed through the EU farm structure survey.

1 Further details on sample size and the survey design are available at: http://webarchive.nationalarchives.gov.uk/20120104121419/http://www.defra.gov.uk/statistics/files/defra-stats-foodfarm-landuselivestock-june-results-methodology.pdf

8.5.2 Pilot survey on vegetable crops in Ghana

Under the Global Strategy to improve Agricultural and Rural Statistics (GSARS), in 2016 a Letter of Agreement was signed by FAO and the Statistics, Research and Information Directorate (SRID) of the Ministry of Food and Agriculture (MOFA) of Ghana for the implementation of a pilot survey test on the estimation of vegetable crops production and area.

The resources allocated to the pilot made it possible to survey two districts: Ada-East in the Greater Accra Region and Keta Municipality in the Volta Region.

Given the absence of an updates sampling frame, survey operations were preceded by the creation of a list frame. The initial idea was to use the lists of farmers belonging to farmer-based organizations (FBOs). However, previous surveys in the country revealed that a very small share of farmers in the country participates in FBOs. Therefore, the list frame was built visiting all neighbourhood of the two districts where vegetable crops were grown.

Survey implementation was organized in the following steps:

- **Sampling design**. A two-stage sampling design was implemented. In the first stage, the PSUs (vegetable growing sites) were selected with probability proportional to size, where the size was given by the number of farms in the site. In the second stage, farmers were selected with equal probability.
- **Listing of vegetable-crop-growing areas**. The first step consisted in listing all horticultural sites. Each site was enumerated with important variables such as geographic position, number of vegetable holders and types of vegetable crops grown. This listing allowed for establishing a sampling frame of vegetable sites that represented the PSU of the survey.
- **Listing of vegetable-growing holdings and identification of the holder**. A complete list of vegetable holdings was obtained for each sample of vegetable crop site. This list was the frame for the selection of SSUs. The information collected were the names of the holders, the crops on each parcel and the specific dates of harvest.
- **First visit**. The main part of the questionnaire was compiled during this visit. The collected data covered socio-demographic characteristics of the holder and his or her household members as well as the economic characteristics of the farm, including labour, land (size and number of plots, number of beds under cultivation, average size of beds by crop, equipment, cost of production, etc.).
- **Yield survey.** A subsample of farms was identified to perform objective production measurements. The production of these farms was harvested and weighed.

TABLE 8.2. VARIABLES COLLECTED TO MEASURE AREA AND PRODUCTION ACCORDING TO EACH CATEGORY OF VEGETABLE CROPS.

CULTURE	Divided into beds		NOT divided into beds	
	Area	Production	Area	Production
Leafy with **single harvest** Lettuce Cabbage Cauliflower	Randomly select **beds** and: Area by inquiry Area by objective measure (GPS) Simple geometric calculations for small beds Number of beds harvested by inquiry over the last 12 months	Randomly select **beds** and: Harvest bed, and weigh Or Count plants Randomly select plants, harvest and weigh Production by farmer inquiry Estimate yield of each harvest by farmer inquiry over the last 12 months	Area by inquiry Area by objective measure (GPS)	Randomly select crop-cutting area Harvest and weigh Production by farmer inquiry Estimate yield of each harvest by farmer inquiry over the last 12 months
Leafy with **staggered harvest** Spinach Mint Parsley Celery Potato leaves	Randomly select **beds** and: Area by farmer inquiry Area by objective measure (GPS) Simple geometric calculations for small beds Number of beds harvested by inquiry over the last 12 months	Randomly select **beds** and: Harvest bed, and weigh Production by farmer inquiry Estimate the number of harvest over the last 12 months by farmer inquiry Estimate yield of each harvest by farmer inquiry over the last 12 months	Area by inquiry Area by objective measure (GPS)	Randomly select crop-cutting area Harvest and weigh Production by farmer inquiry Estimate yield of each harvest by farmer inquiry over the last 12 months
Root vegetables with **single harvest** Carrot Beet Turnip Bulb onion Shallot Leek	Randomly select **beds** and: Area by inquiry Area by objective measure (GPS) Simple geometric calculations for small beds Number of beds harvested by inquiry over the last 12 months	Randomly select **beds** and: Harvest bed, and weigh Or Estimate number of measurement units Harvest randomly selected plants, weight measurement unit Production by farmer inquiry in measurement units Estimate yield of each harvest by farmer inquiry over the last 12 months	Area by inquiry Area by objective measure (GPS)	Randomly select crop-cutting area Harvest and weigh OR Estimate production by farmer inquiry in measurement units Estimate weight of measurement units by farmer inquiry
Fruit vegetables with **single harvest** Pumpkin	Randomly select **beds** and: Area by inquiry Area by objective measure (GPS) Simple geometric calculations for small beds Number of beds harvested by inquiry over the last 12 months	Randomly select **beds** and: Harvest bed, and weigh Or Estimate number of measurement units by farmer inquiry Estimate weight of measurement unit by harvesting, filling, and weighing Estimate yield of each harvest by farmer inquiry over the last 12 months	Area by inquiry Area by objective measure (GPS)	Randomly select **crop-cutting** area Harvest, and weigh Or Estimate number of measurement units by farmer inquiry Estimate weight of measurement unit by harvesting, filling and weighing Estimate yield of each harvest by farmer inquiry over the last 12 months

CULTURE	Divided into beds		NOT divided into beds	
	Area	Production	Area	Production
Fruit vegetables with **staggered harvest** * Cucumber Eggplant Pepper Hot pepper Tomato Okra *Methods are the same regardless of whether fruit vegetables are cultivated in beds.	Randomly select **beds** and: Area by farmer inquiry Area by objective measure (GPS) Simple geometric calculations for small beds Number of beds harvested by inquiry over the last 12 months	Estimate number of measurement units by **farmer inquiry** Estimate weight of measurement unit by harvesting, filling and weighing Estimate the number of harvest over the last 12 months by farmer inquiry Estimate yield of each harvest by farmer inquiry over the last 12 months	Area by inquiry Area by objective measure (GPS)	Estimate number of measurement units by **farmer inquiry** Estimate weight of measurement unit by **crop cutting**, filling and weighing Estimate yield of each harvest by farmer inquiry over the last 12 months

8.5.3 Crop estimation survey on fruits and vegetables in India

Fruits and vegetables account for a substantive part of agricultural production in India. Currently, India is the second largest producer of fruits and vegetables in the world and is the leader in several crops, such as mango, banana, papaya, cashew nut, areca nut, potato and okra (Horticulture Statistics at a Glance, 2017).

Given the importance of fruit and vegetable production, India has always placed great efforts in the collection of data on the sector. In 1988, the National Horticultural Board (NHB) in India launched a "Market Information Service", to generate useful information on price and arrival trends in various markets of the country for important fruits and vegetables. For this purpose, the NHB identified 22 centres and 11 subcentres located in important cities of the country. From these centres, data on wholesale and retail prices and arrivals of important vegetables are collected. The information obtained is disseminated through its monthly publication *Horticulture Information Service*. The publication reveals arrival trends of vegetables of commercial importance in major Indian wholesale markets. To extrapolate a reasonable estimate of vegetable production, besides this administrative source, it is necessary to take into account home consumption from home-grown production, consumption by food-processing establishments, exports and, finally, loss of production, as the crops are perishable in nature.

The Directorate of Economics and Statistics (DES) under the Ministry of Agriculture and Farmers Welfare (MAFW) implemented the Crop Estimation Survey of Fruits and Vegetables (CES-F&V) as a component of the Improvement of Agricultural Statistics Scheme. This survey ran from 1982 to 2014 in 11 Indian States and covered seven fruit crops (apple, mango, citrus, pineapple, grape, banana and guava) and seven vegetables crops (potato, cabbage, cauliflower, onion, tomato, ginger and turmeric). In 2014, the survey scheme has been discontinued and replaced with a new methodology proposed by the Indian Agricultural Statistics Research Institute (IASRI), which is much simpler and easy to implement. This new methodology is currently being test in six states under the CHAMAN project of the Department of Agriculture and Cooperation. The programme has the objective of developing and consolidating the scientific methodology for estimating area and production under horticultural crops. It has two main components:

- remote sensing technology; and
- the sample survey methodology.

For details on the methodology proposed by the CHAMAN project, see Horticulture Statistics at a Glance, 2017.

8.6 MAIN RECOMMENDATIONS TO SURVEY PRACTITIONERS

The implementation of an information system on horticultural crops requires following a number of steps ensuring the sustainability of the process and the quality of the data collected.

- First, the objectives of the data collection exercise – in terms of crops to follow, data to gather and periodicity of data collection – must be clearly stated.
- Second, existing data sources, such as administrative records, must be carefully reviewed.
- Third, the availability of updated sampling frames must be ensured.
- Fourth, the questionnaire used for data collection must be properly field-tested.

The particularity of vegetable crops descends from the fact that a significant part of their production comes from kitchen gardens cultivated mainly for own consumption. The best tool for measuring the area and production of vegetable crops cultivated in kitchen gardens is represented by household surveys, as this type of agricultural production may also represent a significant component of households' income.

However, the greater part of vegetable production is usually carried out by bigger holdings that produce primarily for sales. For this holdings category, two types of data on vegetable crops are particularly interesting:
- Structural data on who is producing and how: the best method to collect these data in a cost-effective manner is to implement an ad hoc module in an annual agricultural survey.
- Data on quantity produced and market prices: generally, this information is useful when collected at a high frequency (on a weekly or monthly basis). A possibility to collect this data consists in surveying markets to collect information on prices and quantities sold.

In the case of ornamental plants, own consumption is normally negligible and production is carried out by specialized holdings. Furthermore, the main sales networks are often not physical markets. Given all these elements, the most cost-effective strategy is to implement an ad hoc module in an agricultural survey. The frequency of data collection should be customized to the specific country context, taking into consideration the importance of vegetable crops production in the country. In many circumstances, an interval of five years between each data collection exercise is sufficient.

9

Integrated survey programmes

9.1. THE NEED FOR AN INTEGRATED SURVEY SYSTEM

In many countries, agricultural surveys are conducted on an ad hoc basis without an overall statistical programme, strategy or links to an MSF. In these cases, it is difficult to integrate data from various surveys for further analysis. In the absence of structural data for the entire agricultural sector that provides the basis for analysing the characteristics of farms, the division of the production of crop and livestock data leaves no opportunity to compare and measure the impact of an action within or between agricultural subsectors.

Household surveys are often conducted in isolation from production surveys or with small sample sizes that cannot disaggregate the rural and farm sectors (World Bank and FAO, 2011). Similarly, data is often collected by subsectors, using different sampling frames and surveys with multiple governmental organizations being involved in its collection of data without coordination from other departments. In some cases, different organizations produce statistics for the same items, with different results.

The Global Strategy to improve Agricultural and Rural Statistics (World Bank, 2011) suggests that an integrated statistical system is capable of resolving many of these problems by reducing the duplication of efforts, preventing the release of conflicting statistics, and ensuring the best use of resources. Methods, concepts and classifications can be standardized and allow for more systematic data collection across sources. These practical advantages of integrated data systems, together with the increasing need for reliable and comparable data in a context of globalization and international concern, point to the need for integrated national statistical systems. This integration will be accomplished by:

1. the development of an MSF for agriculture, including crop statistics;
2. its use within a coordinated data collection program to produce timely and accurate data; and
3. a strategy for effective data dissemination that ensures accessibility.

Indeed, GSARS developed the AGRIS, which proposes a ten-year modular data collection scheme based on an MSF (GSARS, 2015), ensuring data coherence in time and among thematic sets of core variables. The options for building the MSF are left to the discretion of national authorities, as they all inherit different systems. If a decennial agricultural census is operational, it will naturally provide the basis for the elaboration of the MSF. If only a household population census is performed periodically, the inclusion of a targeted agriculture module will enable the derivation of the agriculture master frame for the household sector. In other cases, the construction of an area frame (points or segments) will provide an efficient solution. In all scenarios, the geolocalization of sampling units will be necessary, and multiple frames will generally improve frame quality by merging list and area frames with administrative sources (especially for the non-household sector).

9.2. OVERVIEW OF AGRIS[1]

AGRIS is a farm-based modular multi-year sample survey program that aims to complement other relevant initiatives, such as the World Bank LSMS-ISA, and to scale up these global efforts. As one of the main features of cost-effective methods, AGRIS is designed to help national agencies accelerate the production of quality disaggregated data on the technical, economic, environmental and social dimensions of agricultural holdings. AGRIS builds on the work of the Global Strategy to channel its methodological innovations and achieve real effects on data systems on the ground in a unique manner. AGRIS, being a ten-year integrated survey program, lays the foundations for the creation of an efficient agricultural statistical system. Together with the agricultural census that it complements, a versatile agricultural market information system, and an appropriate use of remote sensing and administrative data, AGRIS is a cornerstone for the establishment of a comprehensive rural information system. AGRIS is designed to serve an integrated national statistical system. It is composed by a Core Module and a series of four Rotating Modules. Each module measures different key aspects of the agricultural sector and is fielded with different frequencies. The AGRIS Core Module is an annual sample farm survey having the main objective of measuring a key set of indicators that are related, in particular, to the volume of agricultural production (crop, livestock, forestry, fishery and aquaculture). In addition, the Core Module measures the key social, economic and technical dimensions of the holding. In addition, a series of Rotating Modules will take place at varying frequencies, when possible based on the samples of the Core Module. These Rotating Modules bring additional knowledge on thematic domains: Economy, Labour, Production Methods and the Environment, and Machinery, Equipment and Assets. AGRIS is mainly conceived for developing countries and is currently being tested in some of them. The recommendations on sampling strategies are issued at two levels. First, commercial farms are separated from households, a fact that favours the design of a two-stage stratified plan for households and a one-stage stratified sampling for commercial holdings. Second, samples should be rotated in time, to limit survey burden and allow for the analysis of longitudinal panel data, which is particularly efficient when observers are interested in estimating the evolution of trends. Finally, the link with SDGs is acknowledged (UN, 2017), and the proposed set of AGRIS Generic Questionnaires will generate basic data for monitoring the relevant SDGs. Among the 232 SDG indicators, AGRIS provides essential and direct information for four SDG indicators and essential but indirect information for another 15 SDG indicators.

1 GSARS. 2018. Handbook on Agricultural Integrated Survey (AGRIS), available at http://gsars.org/wp-content/uploads/2018/05/

9.2.1. The AGRIS cycle

AGRIS is synchronized with the agricultural census and operates over a ten-year cycle. AGRIS seeks to decrease the burden of conducting censuses by scheduling the collection of thematic data over this period. This will contribute to a more regular flow of data, which would be more consistent with the limited capacities currently in place for the production and use of statistics. AGRIS consists of a collection of questions that can be classified into one of two main categories: a core section and a rotating section. The core section (also referred to as the "Core" or "Core Module") is an enhanced production survey that also focuses on a range of different themes, which remain largely the same in each survey round. The rotating section (which comprises several "Rotating Modules") is devoted to specific themes, the implementation frequency of which varies among countries with different agricultural systems and data demand priorities. The following table summarizes a possible module flow for the four recommended Rotating Modules on Economy, Labour, Production Methods and the Environment, and Machinery, Equipment and Assets. The financial and human resources required to sustain and implement such an arrangement is relatively stable over the ten-year cycle, making it a viable set-up for a data producing agency. The flexible, modular nature of AGRIS makes it easy to modify this proposed setting and thus enhance its national relevance and cost-effectiveness. Additional Rotating Modules may also be added to respond to other specific data needs.

TABLE 9.1. RECOMMENDED AGRIS MODULE FLOW.

	Years	1	2	3	4	5	6	7	8	9	10
Core Module	Agricultural holding (AH) Roster	•	•	•	•	•	•	•	•	•	•
	Crop + livestock production	•	•	•	•	•	•	•	•	•	•
	Other key variables	•	•	•	•	•	•	•	•	•	•
Rot. Module 1	Economy	•		•		•		•		•	
Rot. Module 2	Labour		•				•				
Rot. Module 3	Production Methods and the Environment				•				•		
Rot. Module 4	Machinery, Equipment and Assets	•				•					

To enhance respondent recall and provide timely information for market efficiency and decision-making, data collection activities could be conducted several times during the year. This is particularly true for the Core Module in countries with several crop periods. Rotating Modules, in particular the Economy and Labour Modules, could also require several waves of data collection during their years of implementation. Subsampling plans could be used to accommodate budget constraints, while producing more frequent data with different levels of statistical significance. The use of CAPI technologies is recommended to improve data quality and timeliness.

AGRIS covers different technical, economic, environmental and social dimensions of agricultural holdings through its Core Module and its four Rotating Modules.

AGRIS collects sex-disaggregated data on key topics through both the Core and the Rotating Modules. This entails a more refined identification of male- and female-headed households and will help to assess women's contribution to agriculture through labour, and their access to and control of productive assets, resources and services. More details on topics covered can be found in GSARS (2017).

9.2.2. Coverage of SDG indicators

The data generated by AGRIS is meant to inform policy design and implementation, improve market efficiency and support research. AGRIS constitutes an invaluable data source and provides the framework for designing, monitoring and evaluating any agricultural or rural policy or investment. The proposed generic AGRIS questionnaires cover most of the requirements of the farm-level minimum set of core data (MSCD). They also provide basic data for monitoring the relevant indicators for the SDGs, a set of goals adopted by countries on 25 September 2015 to end poverty, protect the planet, and ensure prosperity for all as part of a new sustainable development agenda. Each goal envisages specific targets to be achieved over the next 15 years, and each target has specific indicators. AGRIS provides essential and direct information for the following four SDG indicators:

- 2.3.1: Volume of production per labour unit producers, by sex and indigenous status
- 2.3.2: Average income of small-scale food producers, by sex and indigenous status
- 2.4.1: Proportion of agricultural area under productive and sustainable agriculture
- 5.a.1: (a) Proportion of total agricultural population with ownership or secure rights over agricultural land, by sex; and (b) Share of women among owners or rights-bearers of agricultural land, by type of tenure

AGRIS contributes to the following 15 additional SDG indicators, on the subpopulation of the population associated with agricultural holdings only:

- 1.1.1: Proportion of population below the international poverty line, by sex, age, employment status and geographical location (urban/rural)
- 1.2.1: Proportion of population living below the national poverty line, by sex and age
- 1.2.2: Proportion of men, women and children of all ages living in poverty in all its dimensions according to national definitions
- 1.3.1: Proportion of population covered by social protection floors/systems, by sex, distinguishing children, unemployed persons, older persons, persons with disabilities, pregnant women, newborns, work-injury victims and the poor and the vulnerable
- 1.4.1: Proportion of population living in households with access to basic services
- 1.4.2: Proportion of total adult population with secure tenure rights to land, with legally recognized documentation and who perceive their rights to land as secure, by sex and by type of tenure
- 1.5.1: Number of deaths, missing persons and directly affected persons attributed to disasters per 100 000 population
- 2.5.2: Proportion of local breeds, classified as being at risk, not-at-risk or unknown level of risk of extinction
- 5.5.2: Proportion of women in managerial positions
- 5.b.1: Proportion of individuals who own a mobile telephone, by sex
- 7.1.1: Proportion of population with access to electricity
- 8.7.1: Proportion and number of children aged 5–17 years engaged in child labour, by sex and age
- 9.1.1: Proportion of the rural population who live within 2 km of an all-season road
- 9.c.1: Proportion of population covered by a mobile network, by technology
- 17.8.1 Proportion of individuals using the Internet

9.3. STEPS IDENTIFIED AND THE CONTRIBUTIONS OF THESE GUIDELINES

The present Guidelines contribute significantly to the establishment of an integrated survey system. Some of the main steps of this process are listed in table 9.2 below, along with the corresponding contribution that this publication could make.

Step	How these Guidelines can contribute
Link a country's plan for producing crop statistics to the Global Strategy framework	1) Support the adoption of the first pillar of GSARS in establishing an MSCD on crops, defined in close consultation with stakeholders, to identify current and emerging demands as well as opportunities. 2) Support the adoption of the second pillar of GSARS by improving existing statistical systems, which is achieved by demonstrating and documenting the following: use of existing sampling frames; integration of data from existing sources and in different forms; and employment of benchmarking for quality. 3) Support to the adoption of the third pillar of GSARS by contributing to the sustainability of statistical systems. This is achieved by documenting the evaluation of proposed new collection methods and identification of low-cost approaches and necessary skills.
Identify the core crop statistics required by different stakeholders along with frequency, level of aggregation and precision required	1) Gap analysis identifying high-priority variables and indicators 2) Evaluation of data quality in existing data, including precision as a criterion 3) Level of aggregation linked to production systems' classification and sampling strategy
Identify appropriate data collection methodologies and which agency have the capacity to collect the data	1) Support the identification of the correct data collection strategy in accordance with the main types of crop 2) Support the improvement of methods to measure crop area and production 3) Enhance the inclusion of smallholder farmers in data collection 4) Support the improvement of adopted sampling designs
Identify any capacity building and training required by organizations responsible for implementation	1) Identified tecÚical capacities and skills necessary for various data collection methods 2) Identified training needs associated with new methods, expressed as additional needs due to adoption of new methods 3) Costs of training incorporated into budgets for data collection
Identify resources required to successfully implement data collection	1) Identification of cost-mitigating strategies for data collection practice, enabling lowest-cost implementation

References

Ajayi, O.C. & Waibel, H. 2000. *How accurate are farm size estimates obtained from smallholder farmers in West Africa? Lessons from Côte d'Ivoire*. In Renard, G., Krieg, S., Lawrence, P. & von Oppen, M. (eds), *Farmers and Scientists in a Changing Environment: Assessing Research in West Africa* (pp. 543–550). Margraf Verlag Publishers: Weikersheim, Germany.

Ali, N., Kahn, K., Patel, P. & Gorelick, J. 2009. *Moving the Learning Forward: From Incremental to Transformational Impact on Empowering Smallholder Farmers and Women*. Impact Assessment of the Agricultural Marketing Initiative, West Nile, Uganda. Capstone Report. Publication of George Washington University: Washington, D.C.

Aune, D. & Vogel, F. 2006. *The yield forecasting program of NASS*. NASS Staff Report No. SMB 06-01. Statistical Methods Branch, Estimates Division, National Agricultural Statistics Service, U.S. Department of Agriculture Publication: Washington, D.C.

Badhwar, G.D. 1984. Automatic corn-soybean classification using Landsat MSS data. I. Near-harvest crop proportion estimation. *Remote Sensing of Environment*, 14: 15–29.

Balogoun, I. et al. 2014. Caractérisation Des Systèmes de Production à Base d'anacardier Dans Les Principales Zones de Culture Au Bénin. *Agronomie Africaine* 26(1): 9–22.

Bauer, M.E., Cipra, J.E., Anuta, P.E. & Etheridge, J.A. 1979. Identification and area estimation of agricultural crops by computer classification of Landsat MSS data. *Remote Sensing of Environment*, 8: 75–92.

Beegle, K., de Weerdt, J., Friedman, J. & Gibson, J. 2012. Methods of household consumption measurement through surveys: experimental results from Tanzania. *Journal of Development Economics*, 98: 3–18.

Belward, A.S., de Hoyos, A. 1987. A comparison of supervised maximum-likelihood and decision tree classification for crop cover estimation from multitemporal Landsat MSS data. *International Journal of Remote Sensing*, 8: 229–235.

Berg, E. & Li, J. 2015. *Administrative Data and Statistical Programmes of Developed Countries*. *Improving the Methodology for Using Administrative Data in an Agricultural Statistics System.* Working Paper 3. Global Strategy to Improve Agricultural and Rural Statistics Working Paper: Rome.

Bernardi, M., Delincé, J., Durand, W. & Zhang, N. 2016. *Crop Yield Forecasting: Methodological and Institutional Aspects*. AMIS, FAO, Rome.

Bettio, M., Delincé, J., Bruyas, P. Croi, W. & Eiden, G. 2002. *Area frame surveys: aim, principals and operational surveys*. In Gallego, J. (ed.), *Building Agro Environmental Indicators: focusing on the European area frame survey LUCAS* (pp. 12–28). European Commission Joint Research Centre Publication: Ispra, Italy.

Bizzell, R.M., Hall, F.G., Feiveson, A.H., Bauer, M.E., Davis, B.J., Manila, W.A. & Rice, P.C. 1975. *Results from the crop identification technology assessment for remote sensing (CITARS) projects*. In Cook, J.J. (ed.), *Proceedings of the Tenth International Symposium on Remote Sensing of Environment*, 6–10 October 1975, Environmental Research Institute of Michigan, Ann Arbor, United States of America.

Bogaert, P., Delincé, J. & Kay, S. 2005. Assessing the error of polygonal area measurements: a general formulation with applications to agriculture. *Measurement Science and Technology*, 16: 1170–1178.

Boryan, C.G., Yang, Z., Di, L. & Willis, P. 2017. Developing Crop Specific Area Frame Stratification Based on Geospatial Crop Frequency and Cultivation Data Layers. *Journal of Integrative Agriculture*. 16(2): 312–323.

Bradbury, D. 1996. *Cereals in Europe: A Supplement—Statistical Systems for Measuring Area, Production, and Yield in Austria, Finland, Sweden and Norway*. Eurostat Publication: Luxembourg.

Carletto, C., Deininger, K., Muwonge, J. & Savastano, S. 2010. *Using diaries to improve production statistics: Evidence from Uganda*. Presentation for the Fifth International Conference on Agricultural Statistics, 13–15 October 2010, Kampala.

Brzozowski, M., Crossley, T.F. & Winter, J.K. 2017. A comparison of recall and diary food expenditure data. *Food Policy*, 72: 53–61.

Campbell, N.A., De Boer, E.S. & Hick, P.T. 1987. Some observations on crop profile modeling. *International Journal of Remote Sensing*, 8: 193–201.

Carfagna, E. 1999. *The Use of Area Frame and Remote Sensing for Agricultural Statistics*. Paper prepared for the Conference of European Statisticians, Meeting on Food and Agricultural Statistics in Europe. Geneva, Switzerland.

Carfagna, E. 2007. *A comparison of area frame sample designs for agricultural statistics*. In Bulletin of the International Statistical Institute, Fifty-sixth Session (Lisbon), Proceedings, Special Topics Contributed Paper Meetings (STCPM11), "Agricultural and Rural Statistics" organized by Steiner, M.A. (NASS-USDA), 1470, 22–29 August 2007.

Carfagna, E. & Gallego, F.J. 2005. Using remote sensing for agricultural statistics. *International Statistical Review*, 73(3): 389–404.

Carfagna, E. & Keita, N. 2009. *Use of modern geo-positioning devices in agricultural censuses and surveys*. In Bulletin of the International Statistical Institute, Fifty-seventh Session (Durban, South Africa), Proceedings, Special Topics Contributed Paper Meetings (STCPM22), organized by N. Keita (FAO), 16–22 August 2009.

Carletto, C., Corral, P. & Guelfi, A. 2016. Agricultural commercialization and nutrition revisited: empirical evidence from three African countries. *Food Policy*, 67: 106–118.

Carletto, C., Gourlay, S., Murray, S. & Zezza, A. 2016a. *Land area measurement in household surveys: A guidebook*. World Bank Publication: Washington D.C.

Carletto, C., Gourlay, S., Murray, S. & Zezza, A. 2016b. *Cheaper, faster and more than good enough: Is GPS the new gold standard in land area measurement?* World Bank Policy Research Working Paper no. 7759. World Bank Publication: Washington, D.C.

Carletto, C., Gourlay, S. & Winters, P. 2013. *From guesstimates to GPStimates: Land area measurement and implications for agricultural analysis*. World Bank Policy Research Working Paper no. 6550. World Bank Publication: Washington, D.C.

Carletto, C., Savastano, S. & Zezza, A. 2011. *Fact or artifact: The impact of measurement errors on the farm size–productivity relationship*. World Bank Policy Research Working Paper no. 5908. World Bank Publication: Washington, D.C.

Casley, D.J. & Lury, D.A. 1981. *Data Collection in Developing Countries*. Clarendon Press: Oxford, United Kingdom.

Casley, D.J. & Kumar, K. 1988. *The collection, analysis and use of monitoring and evaluation data*. Publication for the World Bank. Johns Hopkins University Press; Baltimore, MD, USA.

Caudill, C. 1977. *Estimation of Agricultural Production by Sample Surveys, the United States Experience*. In the Proceedings of the 41st Session of the International Statistical Institute, New Delhi.

Central Statistics Organization. 2007. *Manual on Crop Area and Production Statistics*. Central Statistics Organization Publication: New Delhi.

Centre National De Recherche Agronomique (CNRA). 2005. *Bien cultiver le riz irrigué en Côte d'Ivoire*. Accessed on January 2018.

Cochran, W.G. 1977**.** *Sampling Techniques*. John Wiley & Sons: Chichester, UK.

Colwell, J.A. 1999. *Looking into Agricultural Statistics: Experiences from Asia and the Pacific Regional Co-ordination Centre for Research and Development of Coarse Grains, Pulses, Roots and Tuber Crops in the Humid Tropics of Asia and the Pacific*. CGPRT Centre Publication: Bogor, Indonesia.

Coulibaly, K., Vall E., Autfray P. & Sédogo, P.M. 2012. Performance technico-économique des associations maïs/niébé et maïs/mucuna en situation réelle de culture au Burkina Faso: potentiels et contraintes. *Tropicultura*, 30(3): 147–154.

Craig, M. & Atkinson, D. 2013. *A Literature Review of Crop Area Estimation*. FAO Publication: Rome.

Dabalen, A., Etang, A., Hoogeveen, J., Mushi, E., Schipper, Y. & von Engelhardt, J. 2016. *Mobile phone panel surveys in developing countries: a practical guide for microdata collection*. Directions in Development. World Bank Publication: Washington, D.C.

Dalrymple, D.G. 1971. *Survey of Multiple Cropping in Less Developed Nations*. Foreign Agricultural Economic Report No. USDA/FAER-91 Report. Economic Research Service, USDA Publication: Washington, D.C.

David, I.P. 1978. *Non-sampling errors in agricultural surveys: Review, current findings, and suggestions for future research*. Paper presented at the Philippine Statistical Association Annual Conference, 18 June 1978. Manila.

David, I.P. 1989. *The State of Agricultural Statistics in Asia*. Statistical Report Series #12. Economics Office, Asian Development Bank. Asian Development Bank Publication: Manila.

David I.P. 1999. *Ten Years of Statistical System Reengineering in Some Asian Transition Countries*. International Statistical Institute World Congress. Asian Development Bank Publication: Manila.

Davis, B., Di Giuseppe, S. & Zezza, A. 2017. Are African households (not) leaving agriculture? Patterns of households' income sources in rural Sub-Saharan Africa. *Food Policy*, 67: 153–174.

Davies, C. 2009. *Area Frame Design for Agricultural Surveys*. Research and Development Division, National Agricultural Statistics Service, USDA. USDA Publication: Washington, D.C.

Defourny, P. 2017. *Land cover mapping and monitoring*. In Delincé, J. (ed.), Handbook on Remote Sensing for Agricultural Statistics (chapter 2). Global Strategy to improve Agricultural and Rural Statistics (GSARS) Handbook: Rome.

Department for Environment, Food & Rural Affairs (DEFRA). *Horticulture Statistics – metadata*. DEFRA methodologies. [more info needed: place of publication? Website? (is this it: https://www.gov.uk/government/publications/horticultural-statistics-metadata?)]

De Groote, H. & Traoré, O. 2005. The cost of accuracy in crop area estimation. *Agricultural Systems*, 84(1): 21–38.

Deininger K., Carletto, C. Savastano, S. & Muwonge, J. 2012. Can diaries help in improving agricultural production statistics? Evidence from Uganda. *Journal of Development Economics*, 98: 42–50.

Delincé, J. 2001. *A European approach to area frame survey*. In Proceedings of the Conference on Agricultural and Environmental Statistical Applications in Rome (CAESAR), 5–7 June 2001, Vol 2, pp. 463–472.

Desiere, S. & Jolliffe, D. 2018. Land productivity and plot size: is measurement error driving the inverse relationship? *Journal of Development Economics*, 130: 84–98.

Dillon A., Gourlay, S., McGee, K. & Oseni, G. 2016. *Land measurement bias and its empirical implications: Evidence from a validation exercise*. World Bank Policy Research Working Paper No. 7597. World Bank Publication: Washington, D.C.

Diskin, P. 1997. *Agricultural productivity indicators measurement guide*. U.S. Agency for International Development (USAID) Publication: Washington, D.C.

Dorosh, P. & Thurlow, J. 2016. Beyond agriculture versus non-agriculture: decomposing sectoral growth-poverty linkages in five African countries. *World Development*, 109: 440–451.

Ethiopia. Central Statistics Agency. *2011/12 Ethiopian Rural Socioeconomic Survey, Post-planting Questionnaire*.

Erenstein, O., Malik, R.K. & Singh, S. 2007. *Adoption and Impact of Zero-Tillage in the Rice-Wheat Zone of Irrigated Haryana, India*. International Center for Maize and Wheat Improvement (CIMMYT) Publication: New Delhi.

FAO. 1982. *Estimation of Crop Areas and Yields in Agricultural Statistics*. FAO Economic and Social Development Paper. FAO Publication: Rome.

FAO. 1996. *Multiple Frame Agricultural Surveys*. Volume 1. FAO Statistical Development Series 7. FAO Publication: Rome.

FAO. 2006. *Measurement of areas. Simple methods in aquaculture*. FAO training series. FAO Publication: Rome.

FAO. 2014. *Analysis of public expenditure in support of food and agriculture in Malawi, 2006-2013*. FAO Publication: Lilongwe.

FAO. 2017. *World Programme for the Census of Agriculture 2020. Volume 1: Programme, concepts and definition*. FAO Publication : Rome.

Fermont, A.M. & Benson, T. 2011. *Estimating yield of food crops grown by smallholder farmers-A Review in the Uganda context*. Development Strategy and Governance Division, International Food Policy Research Institute (IFPRI) Discussion Strategy. IFPRI Publication: Washington, D.C..

Fermont, A. M., van Asten, P.J.A., Tittonell, P., van Wijk, M.T. & Giller, K.E. 2009. Closing the Cassava Yield Gap: An Analysis from Small-holder Farms in East Africa. *Field Crops Research*, 112: 24–36.

Fielding, W. J. & Riley, J. 1997. *How Big Should On-Farm Trials Be and How Many Plots Should Be Measured?* PLA Notes, 29: 19–22. International Institute of Economic Development, London.

Gallego, F.J. 1995. *Sampling Frame of Square Segments*. Publication of the Office for Publications of the E.C.: Luxembourg.

Gallego, F.J. 2004. Remote sensing and land cover area estimation. *International Journal of Remote Sensing*, 25(15): 3019–3047.

Gallego, F.J. 2007. *Sampling efficiency of the EU point survey LUCAS 2006*. In Proceedings of the 56th International Statistical Institute Session, 22–29 September 2007. Lisbon.

Gallego, F.J. 2015. *Area Sampling Frames for Agricultural and Environmental Statistics: Short guidelines for Developing Countries*. Publication of the Office for Publications of the E.C.: Luxembourg.

Ghana Cocoa Board. 2014. *45th Annual Report and Financial Statements for the Year Ended 30 September* 2014.

Gonzalez-Alonso, F. & Cuevas, J.M. 1993. Remote sensing and agricultural statistics: crop area estimation through regression estimators and confusion matrices. *International Journal of Remote Sensing*, 14(6): 1215–1219.

Gonzalez-Alonso, F., Soria, S.L. & Cuevas-Gozalo, J.M. 1994. Comparing two methodologies for crop area estimation in Spain using Landsat TM images and ground gathered data. *Remote Sensing of Environment*, 35(1): 29–35.

Goswami, S.B., Saxena, A. & Bairagi, G.D. 2012. Remote sensing and GIS based wheat crop acreage estimation in Indore district, M.P. *International Journal of Emerging Technology and Advanced Engineering*, 2(3): 200–203.

Goundan, A. & Magne Domgho, L.V. 2016. *Land area measurement bias: Evidence from West African countries*. MPRA Paper 75476, University Library of Munich: Munich, Germany.

Gourlay S., Kilic, T. & Lobell, D. 2017. *Could the debate be over? errors in farmer-reported production and their implications for the inverse scale-productivity relationship in Uganda*. World Bank Policy Research Working Paper No. 8192. World Bank Publication: Washington, D.C.

GSARS. 2015. *Handbook on Master Sampling Frame for Agricultural Statistics: Frame Development, Sample Design and Estimation*. Global Strategy to Improve Agricultural and Rural Statistics Handbook: Rome.

GSARS. 2018. *AGRIS: Handbook on the Agricultural Integrated Survey*. GSARS Handbook: Rome.

Haack, B. & **Jampoler, S. 1995.** Color composite comparisons for agricultural assessments. *International Journal of Remote Sensing*,16(9): 1589–1598.

Hallum, C.R. & Perry Jr, C.R. 1984. Estimating optimal sampling unit sizes for satellite surveys. *Remote Sensing of Environment*, 14(1–3): 183–196.

Hay, A.M. 1988. The derivation of global estimates from a confusion matrix. *International Journal of Remote Sensing*, 9(8): 1395–1398.

Hay, A.M. 1989. Global estimates from confusion matrices: a reply to Jupp. *International Journal of Remote Sensing*, 10(9): 1571– 1573.

Hixson, M.M., Bauer, M.E. & Scholz, D.K. 1980. *An assessment of Landsat data acquisition history on identification and aerial estimation of corn and soybeans*. LARS Technical Reports. Paper 53. Available at: http://docs.lib.purdue.edu/larstech/5.

Government of India. 2017. *Horticultural Statistics at a Glance 2017*. Horticulture Statistics Division, Department of Agriculture, Cooperation & Farmers Welfare Ministry of Agriculture & Farmers Welfare Publication: New Delhi.

Howard, J., Said, A., Molla, D., Diskin, P. & Bogale, S. 1995. *Toward Increased Domestic Cereals Production in Ethiopia: Using a Commodity Systems Approach to Evaluate Strategic Constraints and Opportunities*. Ministry of Economic Development and Cooperation (Ethiopia) and Michigan state University Food Security Research Project Publication: East Lansing, MI, USA.

Janssen, L.L.F. & Middelkoop, H. 1992. Knowledge-based crop classification of a Landsat Thematic Mapper image. *International Journal of Remote Sensing*, 13(15): 2827–2837.

Jinguji, I. 2015. *Dot sampling method for area sampling*. In the Proceedings of the Expert Meeting on Crop Monitoring for Improved Food Security, 17 February 2014. Vientiane.

Jupp, D.L.B. 1989. The stability of global estimates from confusion matrices. *International Journal of Remote Sensing*, 10(9): 1563–1569.

Kalton G. & Flores-Cervantes, I. 2003. Weighting Methods. *Journal of Official Statistics*, 19(2): 81–97.

Kabambe, V.H. 2011. *Guide cassava production and utilisation in Malawi*. Background paper prepared for the Trustees of Agricultural Promotion Programme.

Kambewa, P. & Nyembe, M. 2008. *Structure and dynamics of Malawi cassava markets*. Background paper prepared for the Cassava Transformation in Southern Africa Start Up Project, Michigan State University.

Keita, N. 2003. *Root and Tuber Crops: Concepts and Methods Recommended by FAO and Operational Issues*. In Proceedings on the Expert Consultation on Root Crop Statistics (chapter 13). Held in Harare, Zimbabwe, 3–6 December 2002.

Keita N., Carfagna, E. & Mu'Ammar, G. 2009. *Issues and guidelines for the emerging use of GPS and PDAs in agricultural statistics in developing countries*. Paper prepared for the Fifth International Conference on Agricultural Statistics (ICAS V), 12–15 October 2010. Kampala.

Keita, N. & Chin, N. 2013. *The place of 'assessment' in current agricultural statistics for developing countries: making best use of available information for timely crop production estimates in the absence of a system of agricultural sample surveys*. Paper presented at the *Sixth International Conference on Agricultural Statistics*, 23-25 October 2013. Rio de Janeiro, Brazil.

Keita N., Ouedraogo, E., Nyamsi, U.E. 2016. *Measuring Area, Yield and Production of Vegetable Crops*. Paper prepared for the ICAS VII Seventh International Conference on Agricultural Statistics I Rome 24-26 October 2016.

Kelly V., Hopkins, J., Reardon, T. & Crawford, E. 1996. *Improving the Measurement and Analysis of African Agricultural Productivity: Promoting Complementarities Between Micro and Macro Data*. Food Security International Development Papers 54055, Michigan State University, Department of Agricultural, Food, and Resource Economics..

Kilic, T., Zezza, A., Carletto, C. & Savastano, S. 2013. *Missing(ness) in action: Selectivity bias in GPS-based land area measurements*. World Bank Policy Research Working Paper No. 6490. World Bank Publication: Washington, D.C.

Kish, L. 1987. *Statistical Design for Research*. John Wiley & Sons: New York, NY, USA.

Kish, L. 1965. *Survey Sampling*. John Wiley & Sons: New York, NY, USA.

Lacroix, E.J. 2003. *Les anacardiers, les noix de cajou et la filière anacarde à Bassila et au Bénin*. GFA Terra Systems : Hamburg, Germany.

Lekasi, J.K., Tanner, J.C., Kimani, S.K. & Harris, J.P.C. 2001. *Managing Manure to Sustain Smallholder Livelihoods in the East African Highlands*. Henry Doubleday Research Association Publication: Ryton on Dunsmore, UK.

Lohr, S.L. 2010. *Sampling: Design and Analysis*. International Edition. Brooks/Cole: Boston, MA, USA.

MacDonald, R.B. & Hall, F.G. 1980. *Global crop forecasting*. *Science*, 208: 670–679.

Malawi Ministry of Agriculture, Irrigation and Water Development (MoAIWD). 2016. *National agriculture policy*. MoAIWD Publication: Lilongwe.

Marshall, M., Husak, G., Michaelsen, J., Funk, C., Pedreros, D. & Adoum A. 2011. Testing a High Resolution Satellite Interpretation Technique for Crop Area Monitoring in Developing Countries. *International Journal of Remote Sensing*, 32(23): 7997–8012.

Magezi-Apuuli, J., Menyha, E., Muwanga-Zake, E.S.K. & Schhøning, P. 2004. *GPS equipment for agricultural statistics surveys – Lessons learned from fieldwork in Uganda*. Paper prepared for the Third International Conference on Agricultural Statistics MEXSAI-ICAS-III, 2-4 November 2004. Cancun, Mexico.

Malawi, National Statistics Office. n.d. *Third Integrated Household Survey 2010/11*, Agriculture Questionnaire.

Martino, L. 2003. *The Agrit system for short-term estimates in agriculture: a project for 2004*. *Polish Seminar: Information Systems in Agriculture*. 9–11 July 2003. Krakow, Poland.

Maselli, F., Conese, C., Zipoli, G. & Pittau, M.A. 1990. Use of error probabilities to improve area estimates based on maximum likelihood classifications. *Remote Sensing of Environment*, 31(2): 155–160.

Mauser, W. 1989. *Agricultural land-use classification in the upper Rhine valley using multi-temporal TM data*. In Proceedings of the Earthnet Pilot Project on Landsat TM Application (ESA-SP - 1102).

McCarthy, N., Kilic, T., de la Fuente, A., & Brubaker, J. 2017. *Shelter from the storm? household-level impacts of, and responses to, the 2015 floods in Malawi*. World Bank Policy Research Working Paper No. 8189. World Bank Publication: Washington, D.C.

Rwanda. Ministry of Agriculture and Animal Resources (MINAGRI). 2009. *Crop Assessment – 2009B Season Final Report*.

Mpyisi, E. 2002. *Estimation of area and production of root and tuber crops in Rwanda*. Paper presented at FAO Expert Consultation on Root Crop Statistics, 2–6 December 2002. Harare.

Moriera, M.A., Chen, S.C. & Batista, G.T. 1986. Wheat-area estimation using digital Landsat MSS data and aerial photographs. *International Journal of Remote Sensing*, 7(9): 1109–1120.

Moylan H., T. Kilic, C. Carletto, and A. Zezza. Forthcoming. *Cassava Production and Productivity Measurement in Household Surveys: A Guidebook for Improving Household Survey Data Collection*. World Bank Publication: Washington, D.C.

Moyo, C.C., Benesi, I.R.M., Chipungu, F.P., Mwale, C.H.L., Sandifolo, V.S. and Mahungu, N.M. 2004. Cassava and sweet potato yield assessment in Malawi. *African Crop Science Journal*, 12(3): 295–303.

Munisse, P., Jensen, B.D., Quilambo, O.A., Andersen, S.B. & Christiansen, J.L. 2012. Watermelon intercropped with cereals under semi-arid conditions: an on-farm study. *Experimental Agriculture* 48(3): 388-398

Murphy, J., Casley, D.J. & Curry, J.J. 1991. *Farmers' Estimations as a Source of Production Data*. World Bank Technical Paper No. 132. World Bank Publication: Washington, D.C.

Muwanga-Zake, E.S.K. 1985. *Sources of possible errors and biases in agricultural statistics in Uganda: A review.* Kampala: Institute of Statistics and Applied Economics, Makerere University.

Malawi. National Statistical Office. 2017. *Fourth Integrated Household Survey (IHS4) 2016/17 survey report*. NSO Publication: Zomba, Malawi.

Nelson, S. & Swindale, A. 2013. *Feed the Future Agricultural Indicators Guide: Guidance on the collection and use of data for selected Feed the Future agricultural indicators*. USAID Publication: Washington, D.C.

Nusser, S. M. & House, C. C. 2009. *Sampling, Data Collection, and Estimation in Agricultural Surveys*. In: Rao, C. R. (eds), *Handbook of Statistics, 29(A): 471-486*. Elsevier B.V.

Palmegiani, G. 2009**.** *Measuring cultivation parcels with GPS: a statistical evidence*. Mimeo, LUISS University: Rome.

Poate, C.D. & Casley, D.J. 1985. *Estimating Crop Production in Development Projects, Methods and Their Limitations*. World Bank Publication: Washington, D.C.

Poate, C.D. 1988. A Review of Methods for Measuring Crop Production from Smallholder Producers. *Experimental Agriculture*, 24: 1–14.

Ray, S. & Neetu. 2017. *Crop area estimation with Remote Sensing*. In: J. Delincé (ed.), Handbook on Remote Sensing for Agricultural Statistics (Chapter 5). Handbook of the Global Strategy to improve Agricultural and Rural Statistics (GSARS): Rome.

Rogers, C.E. & Murfield, D.E. 1965. Validation of Objective Method of Estimating Soybean Yield. *Agricultural Economics Research*, 17(3): 90–91.

Rosales, F.E., Alvarez, J.M. and Vargas, A. 2010. *Guide pratique pour la production de bananes plantain sous haute densité de Plantation : Retours d'expériences d'Amérique latine et des Caraïbes*. Bioversity International Publication: Rome

Rozelle, S. 1991. *Rural Household Data Collection in Developing Countries: Designing Instruments and Methods for Collecting Farm Production Data*. Working Papers in Agricultural Economics No. 91–17. Cornell University: Ithaca, NY, USA.

Ryerson, R.A., Dobbins, R.N. & Thibault, C. 1985. Timely crop area estimates from Landsat. *Photogrammetric Engineering and Remote Sensing*, 51(11): 1735–1743.

Sahoo, P.M., Ahmad, T., Singh, K.N. & Gupta, A.K. 2013. *Study to develop methodology for crop acreage estimation under cloud cover in the satellite imageries*. Project report, Indian Agricultural Statistics Research Institute (IASRI). IASRI Publication: New Delhi.

Sahoo, P.M., Rai, A., Ahmad, T., Singh, R. & Handique, B.K. 2012. Estimation of acreage under paddy crop in Jaintia Hills district of Meghalaya using remote sensing and GIS. *International Journal of Agricultural and Statistical Sciences*, 8(1): 193–202.

Sahoo, P.M., Rai, A., Bathla, H.V.L., Ahmad, T. & Farooqui, S. 2010. *Developing remote sensing based methodology for collection of agricultural statistics in North East hilly region*. Project report, IASRI. IASRI Publication: New Delhi.

Sahoo, P.M., Rai, A., Singh, R., Handique, B.K. & Rao, C.S. 2005. Integrated approach based on remote sensing and GIS for estimation of area under paddy crop in North-Eastern hilly region. *Journal of the Indian Society of Agricultural Statistics*, 59(2): 151–160.

Sapkota ,T.,Jat, M.L., Jat, R.K., Kapoor, P. & Stirling, C. 2016. *Yield Estimation of Food and Non-Food Crops in Smallholder Production Systems*. In Rosenstock, T.S., Rufino, M.C., Butterbach-Bahl, K., Wollenberg, E., Richards, M. (eds), *Methods for Measuring Greenhouse Gas Balances and Evaluating Mitigation Options in Smallholder Agriculture* (chapter 8). Springer Open.

Schøning P., Apuuli, J.B.M., Menyha, E. & Muwanga-Zake, E.S.K. 2005. *Handheld GPS equipment for agricultural statistic surveys: experiments on area measurements done during fieldwork for the Uganda Pilot Census of Agriculture, 2003*. Statistics Norway Publication: Oslo.

Sempungu, A. 2010. *Comparison of the crop card and farmer recall method for cassava and sweet potato yields*. Makerere University, Kampala. (M.Sc. Thesis).

Setiyono, T., Nelson, A. & Holecz, F. 2014. *Remote sensing-based crop yield monitoring and forecasting*. In the Proceedings of the Expert Meeting on Crop Monitoring for Improved Food Security, 17 February 2014. Vientiane.

Shankar Ray, S., Mamatha Neetu, S. & Gupta, S. 2014. *Use of remote sensing in crop forecasting and assessment of impact of natural disasters: operational approaches in India*. In the Proceedings of the Expert Meeting on Crop Monitoring for Improved Food Security, 17 February 2014. Vientiane.

Singh, R. 2003. *Use of Satellite Data and Farmers Eye Estimate for Crop Yield Modeling*. IASRI Publication: New Delhi.

Smale, M.,Diatkite, L., Sidibe, A., Grum, M., Jones, H., Traore, I. & Guindo, H. 2010. The Impact of Participation in Diversity Field Fora on Farmer Management of Millet and Sorghum Varieties in Mali. *African Journal of Agricultural Resource Economics* 4(1): 23–47.

Som, H. 2010. *Characterization of Small Farmers in Asia and the Pacific*. Paper prepared for the Twenty-third Session of the Asia and Pacific Commission for Agricultural Statistics, 26–30 April 2010. Siem Reap, Cambodia.

Srivastava, K.M. (ed.). 2015. *Crop Monitoring for Improved Food Security*. In the Proceedings of the Expert Meeting on Crop Monitoring for Improved Food Security, 17 February 2014. Vientiane.

Ssekiboobo, A.M. 2007. *Practical Problems in the Estimation of Performance Indicators for the Agricultural Sector in Uganda*. Paper presented at the Fourth International Conference on Agricultural Statistics, 22–24 October. Beijing.

Statistics Canada. 2017. *Agricultural Statistics Program Review*. Statistics Canada Publication: Ottawa.

Sud, U.C., Ahmad, T., Gupta, V.K., Chandra, H.,, Sahoo, P.M., Aditya, K., Singh, M. & Biswas, A. 2016. *Measuring Crop Area and Yield Under Pure Stand, Mixed and Continuous Cropping*. Research on improving methods for estimating crop area, yield and production under mixed, repeated and continuous cropping. Global Strategy Working Paper No. 5. GSARS Working Paper: Rome.

Sud U.C., Ahmad, T., Gupta, V.K., Chandra, H., Sahoo, P.M., Aditya, K., Singh, M., Biswas, A. & ICAR-IASRI. 2015. *Gap Analysis and Proposed Methodologies for Estimation of Crop Area and Crop Yield under Mixed and Continuous Cropping*. Research on Improving Methods for Estimating Crop Area, Yield, and Production under Mixed, Repeated and Continuous Cropping. Global Strategy Working Paper No. 4. GSARS Working Paper: Rome.

Sud, U.C., Ahmad, T., Gupta, V.K., Chandra, H., Sahoo, P.M., Aditya, K., Singh, M., Biswas, A. & ICAR-IASRI. 2016. *Synthesis of Literature and Framework*. Research on Improving Methods for Estimating Crop Area, Yield, and Production under Mixed, Repeated and Continuous Cropping. Global Strategy Working Paper No. 5. GSARS Working Paper: Rome.

Sud, U.C., Ahmad, T., Gupta, V.K., Chandra, H., Sahoo, P.M., Aditya, K., Singh, M. & A. Biswas. 2017. *Methodology for Estimation of Crop Area and Crop Yield under Mixed and Continuous Cropping*. Global Strategy Technical Report Series GO-21-2017. GSARS Working Paper: Rome.

Uganda. Ministry of Agriculture and Co-operatives (MAC). 1965. *Report on Uganda census of agriculture*, Vol. I. MAC Publication: Kampala.

Uganda Bureau of Statistics (UBOS). 2002. *Uganda National Household Survey 1999/2000: Report on the Crop Survey Module*. Uganda Bureau of Statistics Publication: Kampala.

Uganda Bureau of Statistics. 2010. *Uganda Census of Agriculture 2008/09 Methodology Report*, Vol. II. Uganda Bureau of Statistics Publication: Kampala.

United States Department of Agriculture (USDA). 1999. *Understanding USDA Crop Forecasts*. USDA Publication: Washington, D.C.

USDA. 2009. *Surveys—Agricultural Yield*. USDA/NASS Publication: Washington, D.C.

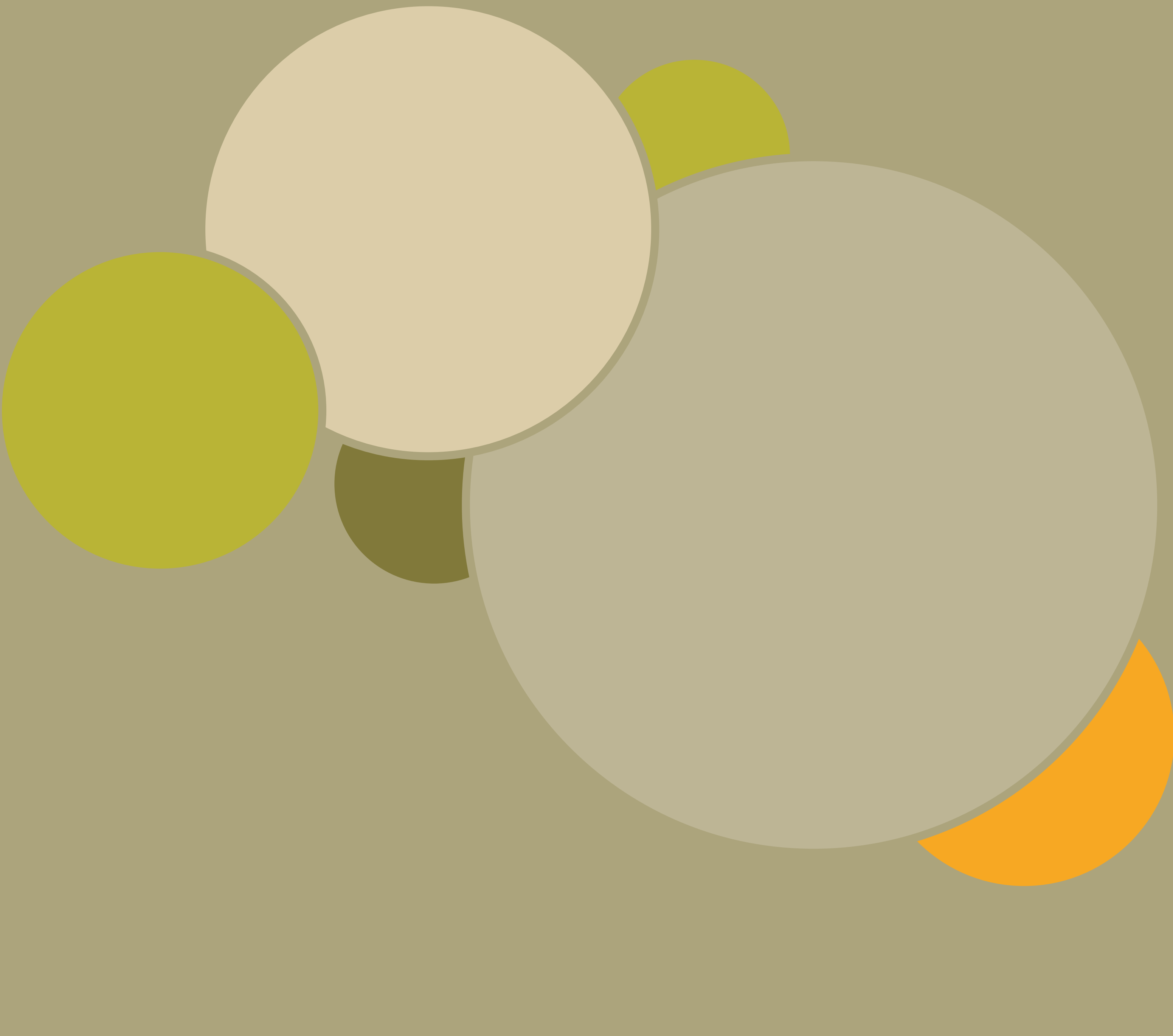
Vibhute, A.D., Nagne, A.D., Gawali, B.W. & Mehrotra, S.C. 2013. Comparative analysis of different supervised classification techniques for spatial land use/land cover pattern mapping using RS and GIS. *International Journal of Scientific and Engineering Research*, 4(7).

Vogel, F. & Bange, G. 1999. *Understanding crop statistics*. U.S. Department of Agriculture, Miscellaneous publication no. 1554. USDA Publication: Washington, D.C.

World Bank. 2011. *Global Strategy to Improve Agricultural and Rural Statistics*. Report No. 56719-GLB. World Bank Publication: Washington, D.C.

Wu, B. & Li, Q. 2004. *Crop area estimation using remote sensing on two-stage stratified sampling*. In Proceedings of the Twentieth Congress of International Society of Photogrammetry and Remote Sensing (ISPRS), Technical Commission, 22–23 July 2004. Istanbul, Turkey.

Zhao, J., Shi, K. & Wei, F. 2007. *Research and application of remote sensing techniques in Chinese agricultural statistics*. Paper presented at the Fourth International Conference on Agricultural Statistics, 22–24 October 2007. Beijing.



Annex 1

Glossary

Agricultural land: total of cropland and permanent meadows and pastures (FAO, WCA 2020). The scope of the definition adopted by the Global Strategy to improve Agricultural and Rural Statistics also includes land for aquaculture production.

Agricultural holding: economic unit of agricultural production under single management, comprising all livestock kept and all land used wholly or partly for agricultural production purposes, without regard to title, legal form, or size (FAO, WCA 2020). If the agricultural holding's principal economic production activity is agricultural production, the agricultural holding is an establishment within the agriculture industry.

Agricultural holder: natural person, group of natural persons or legal person who makes the major decisions regarding resource use and exercises management control over the agricultural holding's operations (FAO, WCA 2020). When the holder is a natural person, he or she is usually the head of the household and the person who makes the day-to-day decisions regarding the operation of the holding. He or she may be called an agricultural operator or a farmer.

Arable land: land used in most years for growing temporary crops (FAO, WCA 2020).

Area Sampling Frame: an area frame is a set of land elements, which may be either points or segments of land. The sampling process may involve single or multiple stages. In most agricultural area frame surveys, the sampling unit is associated with a holding.

Area under cultivation: the total area sown or planted; however, after the harvest, it excludes the ruined area resulting from, for example, natural disasters and calamities. Area under cultivation may be the same or smaller than the sown or planted area.

Bed: smallest piece of area where vegetables are planted, with paths all around. Paths are necessary to water and weed without having to tread on the bed. When the soil is poor or does not drain well, beds can be built and raised. The sum of bed areas constitutes the reference area for estimating the production, yield and area harvested of vegetable crops.

Census: statistical collection in which all units are enumerated (large sample-based collections may also be referred to as "sample censuses"). This also means that the relevant data collection includes the entire target population.

Census coverage: the geographical regions of a country covered by census activities. Countries may omit certain areas – for example urban areas, remote areas or those affected by security problems – for operational reasons (FAO, WCA 2020). Holdings of less than a given cut-off point in terms of land area or other size-related variables (for example, household plots) may also be excluded.

Census of agriculture: statistical operation for collecting, processing and disseminating data on the structure and output of agriculture, covering the whole or a significant part of the country (FAO, WCA 2020).

Census reference day: a point in time used for data collection on livestock numbers and inventory items (FAO, WCA 2020).

Census reference year: period of twelve months, either a calendar year or an agricultural year, generally encompassing the various time reference dates or periods of data collection for individual census items (FAO, WCA 2020).

Census scope: the types of agricultural production activities included in the agricultural census. The scope of the agricultural production industry can be interpreted very broadly, to cover not only crop and livestock production activities but also forestry and fisheries production activities, as well as other food and agriculture-related activities (FAO, WCA 2020).

Closed segment: a method for defining a reporting unit when the sampling unit is a segment of land selected from an area sampling frame. The reporting unit is a tract of land within the segment boundaries comprising all or part of a holding. The data are collected only for the land within the segment boundaries.

Cluster sampling: a sampling method that harnesses the physical proximity of population elements so that frame development and data collection costs are reduced. The population is partitioned into primary sampling units (PSUs) or clusters or group of elements from which a number of clusters are sampled. The sampled PSUs can then be enumerated wholly or elements can be drawn from the sampled clusters. (Kish, 1989; p. 74).

Crop area: this can be defined as "the horizontal projection of a particular extent of earth's surface" which corresponds to the area shown on cadastral maps that has crops. This definition takes care of crop areas in the plains and in hilly regions. It also ensures that the total area is equal to the sum of the component area, which is not the case when an area is measured on slopes. The area planted may not be equal to the area harvested and, therefore, the concept of crop area must be refined into sown or planted area and harvested area. Moreover, some crops, such as cassava, may be grown as an insurance measure and are only fully harvested during a drought or a food shortage. In any of the above circumstances, the definition of crop area that is used has a large influence on area, yield and production estimates.

Cropland: total arable land and land under permanent crops (FAO, WCA 2020).

Crop yield: The concept of crop yield is generally used to represent the average amount of produce obtained per unit of the crop area, while the concept of production covers the total amount produced (FAO, 1982). Regulation (EC) No. 543/2009 on crops statistics defines crop yield as the harvested production per unit area under cultivation. In case of tree crops, the concept of yield covers the average amount of produce per tree and the production is calculated as the product of the average yield per tree and the number of producing trees. Some of the commonly used methods for measuring crop yield are crop cut and farmers' appraisal. The three main concepts of crop yield used by many countries are described below (Fermont and Benson, 2011):

- **Biological yield or gross yield**: the yield obtained before any loss occurs during and after harvest;
- **Harvested yield**: the biological yield minus harvest losses;
- **Economic yield**: the quantity that the farmer can use after postharvest losses that may occur during cleaning, threshing, winnowing and drying (Casley and Kumar, 1988; Keita, 2003).

Complete enumeration: collection of data from all units, rather than from only a sample of units.

Computer-Assisted Personal Interview (CAPI): an interview method in which the enumerator conducts an interview with the respondent using an electronic questionnaire or mobile devices – such as tablets, laptops, or smartphones – which the enumerator uses to record the responses. When connected to the Internet or a telephone network, the data captured can be transferred and centralized immediately after the interview. The results can be directly arranged in a format that can be read by statistical analysis programs, an aspect that substantially reduces data processing time.

Continuous planting or harvesting: repeated planting and harvesting of crops at particular intervals of time in an agricultural year. The practice of continuous planting is very common in many African countries. For example, farmers in Uganda plant and harvest crops throughout the year.

Counting units: PSUs in an area frame. Units are randomly selected, and then divided into sample segments for data collection purposes.

Crop area: the physical areas of land on which crops are grown (often referred to as "net cropped area"). The sum of the areas of all temporary crops grown (gross cropped area) may be greater than the net cropped area due to successive cropping (WCA 2020). In hilly regions with abrupt slopes (having an incline greater than 20 percent), the crop area should not be the physical area measured on the slope (the inclined plane), but rather its projection on a horizontal plane (FAO, 1982; para 28).

Crop rotation: crop rotation is a method of farming in which a number of different plants are grown one after the other on a field so that the soil stays healthy and fertile, that is, able to produce crops.

Cut-off threshold: the minimum size limit for inclusion of units in the census (FAO, WCA 2020).

Developed area: the total area used to produce one vegetable. For example, a holding produce radishes each month during six months on a bed of 200 m^2. The part of agricultural area utilized for radishes is 200 m^2; however, the developed area to take into account in the calculation of radishes yield is $200\ m^2 \times 6 = 1\ 200\ m^2$. This developed area can be the sum of several beds, each one used once or more.

Frame: the set of source materials from which the sample is selected (UN, 2005). It is the basis for identifying all statistical units to be enumerated in a statistical collection.

Global Navigation Satellite System (GNSS): satellite navigation system used to determine an object's ground position. A GNSS device enables identification of the geographic position of a point on the Earth's surface by longitude and latitude. It can enable the geo-referencing of the holding, the household and the land to the appropriate administrative areas. As the Global Positioning System (GPS; see below) is the most popular GNSS system, GNSS are often called GPS.

Global Positioning System (GPS): the most popular GNSS in existence today that provides geographic location and time information to a receiver, such as an application in a smartphone anywhere on or near the Earth where there is an unobstructed line of sight to four or more GPS satellites. It is a space-based radio-navigation system owned by the United States Government and operated by the United States Air Force.

Greenhouses: part of vegetables for consumption in fresh and ornamental plants are growing under greenhouses, heated or not but always irrigated. A greenhouse is a fixed building where temperature and aeration are managed following the plants' needs.

Growing season: the growing season in a particular country or area is the period in each year when the weather and temperature are right for plants and crops to grow.

Growth cycle: for each plant, it is the period of growth between sowing or planting and harvesting. There are several vegetative phases during the growing cycle, such as germination and flowering. Generally, harvest is carried out during the developing phase with a maximum yield.

Harvested area: FAO (Martinez et al., 2015) and Regulation (EC) No. 543/2009 on crop statistics defines harvested areas as the part of the sown or planted area that is harvested. The harvested area may, therefore, be equal to or less than the planted area. It serves as an important basis for obtaining a reliable and accurate yield and production estimates.

Head of household: the head of household is the member of the household who generally runs the affairs of the household, and is looked upon by the other members of the household as the main decision maker.

Holder: see agricultural holder.

Holding: see agricultural holding.

Household: the concept of household is based on the arrangements made by persons, individually or in groups, for providing themselves with food or other essentials for living. A household may be either (a) a one-person household, that is, a person who makes provision for his or her own food or other essentials for living without combining with any other person to form part of a multiperson household or (b) a multiperson household, that is, a group of two or more persons living together who make common provision for food or other essentials for living. The persons in the group may pool their incomes and may, to a greater or lesser extent, share a common budget; they may be related or unrelated persons or constitute a combination of persons both related and unrelated. A household may be located in a housing unit or in a set of collective living quarters, such as a boarding house, a hotel or a camp, or may comprise the administrative personnel in an institution. The household may also be homeless (UN, 1998, para. 1.324; FAO, 2020).

Housing census: the overall process of planning, collecting, compiling, evaluating, disseminating and analysing statistical data relating to the number and condition of the housing units and facilities available to the households, concerning, at a specified time, all living quarters and occupants thereof within a country or a well-delimited part of a country.

Hydroponic: hydroponic plants are grown in greenhouses without any soil and are constantly fed by a nutrient solution that is pumped from a tank (reservoir) and is, most often, continually recycled throughout the system. Vegetables such as lettuce, tomatoes, peppers, beans, and cucumbers are all suitable hydroponic crops. Hydroponic crops mostly include food crops rather than ornamental crops such as flowers, although many hydroponic growers are starting to grow flowers in addition to edible crops.

Field: a piece of land in a parcel that is separate from the rest of the parcel by easily recognizable demarcation lines, such as paths, cadastral boundaries, fences, waterways or hedges (FAO, WCA 2020). A field may consist of one or more "plots"; a plot is a part or the whole of a field on which a specific crop or crop mixture is cultivated.

Kitchen garden: the area devoted to the cultivation of agricultural products intended exclusively or at least for 80 percent or 90 percent for own consumption by the household and its relatives. Following these rules, a kitchen garden can be located in any kind of household in or outside of the agricultural sector. In some countries, kitchen gardens' production is considered as part of household production and is estimated from household surveys. In this case, it is important to avoid any double count with agricultural survey sources.

Land cover: the observed biophysical coverage of land. Examples are cropland, woodland, shrub land, grassland, artificial land, bare land, water areas or wetlands. A single location may comprise multiple land covers (for example, crops under tree cover).

Land use: the socio-economic use that is made of land, such as for agriculture, forestry, aquaculture and fishing, mining and quarrying, industry, commerce, residential or unused. A single location may comprise multiple land uses (for example, woodland used for feeding cattle and forestry). Land used for agriculture: total of "agricultural land" and "land under farm buildings and farmyards" (FAO, WCA 2020)

List sampling frame: in these Guidelines, a list frame is a list of farms or households obtained from agricultural or population censuses or administrative data. It is to be noted that the ultimate sampling units are the lists of names of holders or households.

Master Sampling Frame (MSF): a frame that enables selection of different samples (including from different sampling designs) for specific purposes: agricultural surveys, household surveys and farm management surveys. The MSF's distinguishing feature is that it enables samples to be drawn for several different surveys or different rounds of the same survey, which makes it possible to avoid building an ad hoc frame for each survey. In the context of the Global Strategy, an MSF is a frame or a combination of frames that covers the population of interest in its entirety, and that enables the linkage of the farm as an economic unit to the household as a social unit, and both of these to the land as an environmental unit. MSFs are designed to enable the integration of agriculture into national statistical systems by establishing a closer link between results from different statistical processes and units.

Microdata: data on the characteristics of the units of a population, such as individuals, households or establishments, which are collected by means of censuses, surveys or experiments (United States Bureau of the Census, 1998, Section 3.4.4).

Minimum set of core data: set of data that each country should provide to facilitate international comparison. The minimum set is defined by the Global Strategy's first pillar.

Mixed crops: two or more different temporary and permanent crops grown simultaneously in the same field or plot (FAO, 1982). Each crop is referred to as an associated crop.

Mixed intercropping: growing two or more crops simultaneously on the same piece of land with no distinct row arrangement is called mixed intercropping. This practice is more commonly applied in traditional and subsistence farming in many developing countries. For example, in some parts of India, large numbers of crops are sown in mixed intercrop arrangements.

Mono-cropping: the practice of growing only one crop on a piece of land year after year is termed as mono-cropping, such as growing only winter season crops in a dryland area. This may be due to climatic or socio-economic conditions or the result of specialization of a farmer in growing a particular crop. For instance, groundnut, cotton and sorghum are grown year after year due to scarce rainfall in different countries. This term can be better described as continuous cropping or continuous mono-cropping.

Multiple frame survey: a sample survey that uses multiple sampling frames. In the context of agriculture, this includes the joint use of area and list sampling frames. The frames are usually not independent of one another; some of the frame units in one frame may be present in another frame.

Multi-stage sampling: a sampling method that, for agriculture, uses large geographical areas or clusters as the first stage. The final sample frame is then developed only within the selected clusters in one or more stages of sampling. In a two-stage sampling design, the clusters are subsampled and the secondary units sampled are the reporting units. In a three-stage sampling design, sampled selected units are subsampled again. Generally, a multistage sampling design is the subsampling (in two or more stages) of primary sampling units (clusters).

Multiphase sampling: in this type of sampling, a large sample is selected in the first phase; from this, subsamples are selected in a second phase. If a given stratification approach is too expensive to be applied to the entire population, it can be applied only to the sample obtained in the first phase (incomplete stratification). The procedure is often used for area frames of points.

Net cropped area: physical area of land on which temporary crops are grown (FAO, WCA 2020).

Non-sampling error: any error that may arise in the entire survey process (from frame development to data analysis) that is systematic or random and is not related to a random error in sampling. These errors include over- or undercoverage of the sample frame, errors resulting from poorly worded questionnaires, etc.

Open segment: a method for defining a reporting unit when the sampling unit is a segment of land selected from an area sampling frame. The reporting unit depends on the location of the headquarters or household of the holder. If it falls within a sample segment, data are collected for the holding's entire operation, regardless of whether it is included in the segment. No data are collected for holdings with land within the segment but whose headquarters are outside the segment.

Open-air field: the majority of vegetable production and part of the production of ornamental plants are carried out on soil without any cover, but generally with irrigation. Vegetables in open-air fields (for example, for industry or market gardening) or ornamental plants can be carried out under low shelters. This kind of cover can be easily transferred from fields to fields.

Ortho-photograph: a photograph that has been modified such that its geometry corresponds to the geometry of a cartographic projection. Traditionally, the ortho-correction process was applied to aerial photographs, by means of analogic procedures; these, however, have been completely replaced with digital procedures. Ortho-correction is also essential to the analysis of satellite images.

Parcel: any piece of land, of one land tenure type, that is entirely surrounded by other land, water, road, forest or other features not forming part of the holding or forming part of the holding under a different land tenure type. A parcel may consist of one or more fields adjacent to each other (FAO, WCA 2020). The concept of "parcel" used in agricultural censuses and surveys may not be consistent with that used in cadastral work. The reference period is a point of time, usually the day of enumeration.

Permanent crops: crops having a growing cycle greater than one year (FAO, WCA 2020). They are sown or planted once and need not be replanted after each annual harvest.

Point sampling: the final sampling unit is a point. The reporting unit is the holding associated with the land that covers the point.

Population: any finite or infinite collection of individuals (ISI Dictionary of Statistical Terms). A population, or target population, is the finite set of all elementary units (sampling units) about which information is sought. Depending on the survey's goals, the elementary units – or simply the elements – of a population may have different forms. Three typical types of elements are holdings or farms, holders or farmers, and households or dwellings. In addition to the nature of its elements, defining a population requires identification of a place and a point in time. Hence, examples of populations are the set of all holders of a province in 2014 or the set of all households of a region in a given year.

Population census: the total process of planning, collecting, compiling, evaluating, disseminating and analysing demographic, economic and social data at the smallest geographical level pertaining, at a specified time, to all persons in a country or in a well-delimited part of a country (http://unstats.un.org/unsd/statcom/doc15/BG-Censuses.pdf).

Primary sampling unit: see cluster.

Primary crops: crops that come directly from the land without having undergone any real processing apart from cleaning (FAO, 2011). They can be further divided into temporary and permanent crops.

Probability proportional to size: a sampling procedure whereby the probability of selection of each unit in the universe is proportional to the size of some known relevant variable (OECD, 2004). Measures of size, such as land area or number of animals, associated with each holding are used to select sampling units with probabilities proportional to size. These measures are usually obtained from a previous data collection.

Production area: in connection with permanent crops, production area means the area that can potentially be harvested in the reference harvest year. It excludes all non-producing areas, such as new plantations that have not yet started to produce crops.

Quality assurance: this covers measurements of the relevance, accuracy, reliability, timeliness and punctuality, accessibility and clarity, comparability and coherence of the data (FAO, WCA 2020).

Reference group: the group of holdings to be tabulated for the item. For example, the item "area irrigated" is only meaningful for land holdings (FAO, WCA 2020).

Replicated sampling: a sampling method used to simplify the estimation of sampling errors or to facilitate rotating panel surveys. This sampling procedure selects m independent samples of equal size n/m instead of drawing one large sample of size n. It enables variance estimation in systematic sampling.

Rural household: household living in areas designated as rural areas, which are usually defined as such by the population census (FAO, WCA 2020).

Sample enumeration: a procedure that consists of sampling the whole or part of the target population, as opposed to the complete enumeration that occurs in censuses.

Sample survey: the collection of data from a sample of units, rather than from all the units (as occurs in a census).

Sampling error: any random sampling method can produce several different samples that can produce a set of statistics. The sampling error is the variability in the results that are obtained from the different samples. Suppose that N = 10 farms and a random sample of n =2 is selected. There are 45 different combinations of two that can be selected from the ten farms resulting in different sample estimates. The standard error is a measure of the variability between these different sample estimates.

Sampling frame: see frame.

Segment: final land unit selected from an area sampling frame.

Single harvest crop: the whole crop production is harvested at the same time and a new plantation must be done to obtain a new production.

Single-stage sampling: sampling scheme in which the sample is selected directly from a list of units covered by the survey.

Sole crop: a crop grown in pure stand.

Statistical unit: the basic unit for which data are collected (FAO, WCA 2020).

Staggered crop: staggered planting is growing the same vegetable, but planting the seeds on different dates throughout the season so that fresh vegetables can be produced for a longer period. This technique is mainly used in small-scale holdings.

Structural data: data on the basic organizational structure of agricultural holdings that do not change quickly over time. Examples are farm size and land use.

Successive cropping: planting and harvesting either the same crop or different crops more than once in the same field during the agricultural year (one crop is planted after the other crop is harvested) is termed as successive cropping.

Temporary crops: crops that are both sown and harvested during the same agricultural year, sometimes more than once.

Weighted segment estimator: a method for defining a reporting unit when the sampling unit is a segment of land selected from an area sampling frame. The reporting unit is all the land operated by every holding that also has land within the sample segment. The estimator is based on the ratio of the holder’s land in the segment to the land area in the entire operation.

Annex 2
Country experiences

Brazil

The Ministry of Agriculture, Livestock and Supply, through the National Food Supply Company (CONAB), systematically carries out assessment of agricultural crops to quantify and monitor Brazilian production. To obtain information about the estimated area of major crops, CONAB uses satellite imagery, aerial photography, and georeferenced information from mapping cultivated areas in the main producing states. This activity began with the launch of the GeoSafras Project, which aims to improve the methodology of the crop forecasting in the country through the development of technologies related to remote sensing, satellite positioning, geographic information systems and statistical, spectral and agrometeorological models to be applied in the estimates of area and yield.

By providing accurate crop area information, mappings are also used in the spectral and agro-meteorological monitoring of crops by monitoring the meteorological conditions, such as rainfall, soil moisture and temperature, and of vegetation indices calculated from satellite images. Historical data of agrometeorological and spectral parameters in the mapped areas in each crop year are also utilized in the development, calibration, and application of models or systems to estimate productivity.

Another agency, the Brazilian Institute of Geography and Statistics (IBGE) designed and administered the System of Integrated Household Surveys based on a master sample, which is used by all household surveys. IBGE adopted this same concept to address the country's need of quality and cost-effective agricultural statistics; developing an MSF for agricultural surveys, by initiating a National System of Agricultural Establishment Sampling Surveys (Sistema Nacional de Pesquisas por Amostragem de Estabelecimentos Agropecuarios – SNPA in Portuguese).

The MSF was primarily based on the 2006 Agricultural Census, but also uses information from the 2010 Population Census. The former listed 5.2 million agricultural holdings, while the latter listed only 2.6 million. However, building the area frame for this MSF turned out to be difficult despite several attempts as it was not possible to identify a large number of units that were present in more than one register.

To overcome the difficulty mentioned above and the discrepancies between the data of the 2006 Agricultural Census and the 2010 Population Census, IBGE studied alternative methodologies for building the MSF for implementing the SNPA. The first alternative is to use administrative records and build the MSF by developing a list frame. In this scenario, the SNPA's sampling frame would be a list of rural producers. Second, the variable totals must be distributed equally among the new enumeration areas. This methodology uses the results of the new Project of Land Use and Coverage developed by IBGE's Geosciences Directorate. Both alternatives require surveys, studies and field experiments to evaluate their viability and efficiency; therefore, the concept of master frame for agriculture is not yet operational in Brazil.

IBGE has experience with building a sampling frame for agriculture. Then, it was decided to follow an area sampling frame design. Under the final design, stratification was executed according to land use and systematically selected in a single stage, with equal probability and with no substitutions. The stratification was established according to the rate of cultivated land or by the predominance of crops, divided into Counting Units (CU) or PSUs; these were subdivided in Area Segments, the survey's ultimate sampling units (USU). Additionally, this area frame was complemented by a list frame, that consisted of a relatively small number of Special Holdings based on the 1985 agricultural census information and that accounted for a large percentage of the total variable. This was updated every year, making it a multiple frame.

The survey was successful for providing acceptable results. However, it was ultimately discontinued because of several reasons: shortage of financial resources, personnel and institutional support, lack of national coverage, and lack of timeliness in providing results. These problems emphasize the need for guidelines on constructing an MSF that considers factors beyond sampling and frame construction, such as measurement methods, timeliness, and resource equipment.

India

A large part of the Indian population depends primarily on agriculture for food and livelihood security. To accommodate this vast majority, policies and programmes relating to agriculture are structured on the basis of reliable agricultural statistics. Therefore, the Directorate of Economics and Statistics (DES), in cooperation with other State institutions, annually carries out the General Crop Estimation Surveys (GCES), commonly known as Crop Estimation Surveys (CES), to address this demand for relevant information. The main goal is to obtain fairly precise estimates of the average yield of principal food and non-food crops for each State and Union Territory (UT) and for the lower administrative units such as Districts, Blocks, etc. The CES is conducted in 29 States and UTs, which serve as the domains of the survey, to gather a comprehensive data profile on 52 food crops and 16 non-food crops. It collects information about the experimental crop and its condition in the field, sowing methods, crop rotation, cropping pattern, crop condition on density, and input details such as seeds, fertilizers, pesticides and irrigation.

The estimates of crop production are obtained by multiplying area estimates by corresponding yield estimates. Crop area statistics are obtained by dividing the States into three broad categories. The first category covers States and UTs which follow the land record system. These account for approximately 86 percent of reporting area and are covered under the Timely Reporting Scheme (TRS). The second category covers states where data are collected on the basis of sample surveys. Last, the third category covers the places where there is no functioning reporting agency; data collection is entrusted to the village headmen. On the other hand, yield estimates are obtained through analysis of Crop Cutting Experiments (CCEs) conducted under the CES. At the stratum level, the average yield is obtained as a simple arithmetic mean of the net plot yields of all experiments. At the district level, average yield of the dry marketable produce per hectare is calculated as a weighted average of the stratum level estimate, where weight is taken from the corresponding crop areas.

The traditional method of estimating crop production by means of eye-estimation has been gradually replaced by the use of CCEs. Under the CES design, the domains are stratified according to the subdistricts: Tehsils, Taluks/ Revenue Inspector Circle, Community Development Blocks, Anchals, etc. Within each stratum, a list of revenue or administrative villages is used as the sampling frame. During the first stage of the selection process, an SRS of villages is chosen without replacement. Within each chosen village, the survey numbers or fields, which serve as the SSUs, are selected from a village register of land record known as the "Khasra" register. Generally, two survey numbers or fields growing the experimental crop are selected for conducting crop experiments. Finally, experimental plots of a specified shape and size are chosen as the ultimate units of sampling (USUs) within each sampled SSU. These experimental plots are portioned across districts and then to strata within a district. The allocation scheme utilized is by proportion according to area under a specified crop or variety (PPS) within the district or strata within the district.

Periodical forecasts are released by the DES, Ministry of Agriculture (DESMOA) as the Government needs advance estimates of production to make decisions in pricing, distribution, export and import, etc. These forecasts are usually based on preliminary eye-estimation and analysis of the priority enumeration of areas and CCEs. When crops are sown as a single crop, the official yield rates as released by the Ministry of Agriculture are statistically the same as the estimates obtained through the CES. In the case of crops that are sown as mixed, and the procedure of apportionment varies from State to State, the two yield estimates usually differ.

Lesotho

The Bureau of Statistics (BOS) conducts agricultural production surveys (APS) annually. Every ten years, the BOS also carries out the agricultural census to generate data on variables that do not change significantly over time and to serve as an indispensable base for assessing reliability of current agricultural statistics. Both the APS and the agricultural census concentrate on the production of both livestock and crops in the rural parts of the country.

Lesotho has integrated non-agricultural and agricultural surveys that share the same MSF. The current MSF was constructed from the 2006 Population and Housing Census; therefore, the EAs are well-defined and cover the country's total area, with no duplication nor overlapping of units; this ensures complete coverage. The country is divided into ten districts, which are regarded as domains of study for all BOS surveys. Within the districts are four agro-ecological zones, each of which experience different climatic conditions. For this reason, these zones are deemed to be strata.

The APS employed a stratified multistage cluster sampling design in selecting the survey sample. In the rural areas, where an EA is delineated to have from 100 to 150 households, the first stage of the three-stage sample design was the selection of PSUs, which are constructed by combining adjacent EAs from the population and housing census. This selection is done with probability proportional to size (PPS) where the number of households are taken as a measure of size. In the urban areas, where the range is from approximately 80 to 100 households, EAs are used as PSUs. All households in the chosen PSUs are listed and classified with categories such as "operating at least one field" or "presence of livestock". Then, from each of the selected PSUs, a sample of farming households is obtained by systematic sampling technique.

For the estimation of crop yield in each sample PSU, a maximum of 15 fields under each principal crop are selected, with equal probabilities for each main crop. These fields constitute third-stage sampling units. The selected PSUs are also subsampled for the continuous multipurpose household survey, which is maintained in parallel with the annual agricultural survey as their data collection is done in the same areas.

Philippines

One of the surveys conducted by the Philippine Statistics Authority (PSA) is the quarterly Rice and Corn Production Survey (RCPS), now known as the Palay and Corn Production Survey (PCPS). This survey was previously conducted by the Bureau of Agricultural Statistics (BAS), before the BAS, together with the National Statistics Office, National Statistics Coordination Board and the Bureau of Labor and Employment Statistics, were merged into the PSA. The survey's main objective is to generate estimates and forecasts on palay (rice) or corn area, production and yield to provide data inputs for policy and programmes on rice and corn.

The PCPS is carried out on a quarterly basis, with the months January, April, July, and October as the reference periods. It gathers information in the first ten days of the quarter by using questionnaires from all sample farming households in sample barangays (villages) in all provinces, except Batanes but including Zamboanga and Davao cities, on the following:

i. area planted or harvested and production by ecosystem, crop type and by seed class;
ii. monthly distribution of production and area harvested;
iii. farm household disposition/utilization of production;
iv. area with standing crop;
v. planting intentions indicator;
vi. use of seeds, fertilizers, and pesticides; and
vii. awareness and availment of Department of Agriculture (DA) Rice and Corn Programs.

The sampling frame for PCPS was based on the 1991 Census of Agriculture and Fisheries, modified such that the provinces of Laguna, Isabela and Bukidnon conducted traditional complete-enumeration strategy. The list of barangays with the corresponding total palay farm area devoted to palay or corn served as the sampling frame for the selection of sample barangays, while the list of farming households in the selected barangays served as sampling frame for the selection of sample households. The list of farming households in the sample barangays was updated in 2007 from which a new set of sample households was drawn. The survey questionnaire was also modified in December 2007 for the January 2008 round. Another round of updating activity was done in the third quarter of 2011.

A replicated two-stage stratified sampling design was used with the barangay as the PSU and farming household as the SSU. For each province, barangays were stratified into ten strata, based on their palay or corn area. The tenth stratum which comprised the largest barangay in the municipalities as identified from the 1991 Census of Agriculture and Fisheries are the certainty barangays. The selection of the remaining barangays for the first to ninth strata used PPS. Four replicates (four independent sets of sample barangays) per stratum were drawn. From the selected barangays, 4 to 25 households (SSUs) were drawn through systematic sampling.

Before results are produced, the PSA implemented a stringent process of quality control measures. Rounds of reviews of the survey instruments were made before the field operations. Before the results are summarized, field data editing, which includes item-by-item checks on the consistency, completeness and acceptability of the data, is done during and after data collection. Another layer of consistency and completeness checks is done during electronic data processing. Once the output table is generated, a series of reviews on the results follows before the data are submitted to the units concerned. Completion is reached when the estimates generated are affirmed at the end of the National Data Review and made part of the Report on the Performance of Philippine Agriculture.

United States: National Resources Inventory (Area Frame)

In compliance with its mandate under the Rural Development Act of 1972, the Soil and Water Resources Conservation Act (RCA) of 1977 and other supporting legislation, the Natural Resources Conservation Service (NRCS) under the U.S. Department of Agriculture (USDA), in collaboration with Iowa State University – Center for Survey Statistics and Methodology (ISU-CSSM), conducts the national resources inventory (NRI) every five years; periodic NRIs were conducted in 1982, 1987, 1992 and 1997. The permanent sampling frame of the NRI was constructed for the purpose of providing updated and credible information on the status and condition of soil, water and related resources of the country's non-federal lands. Furthermore, the trends of these resources are essential to estimating their changes.

Since 2000, NRI data have been gathered annually, although major releases of these data continue to be reported at five-year intervals. This poses statistical and practical advantages. In addition to better efficiency in conducting the survey and balancing of resources, annualization also makes it easier for the NRI to respond to changing resource issues.

The universe of interest for the NRI survey consists of all surface area (land and water) of the United States of America, including all 50 states, Puerto Rico, the U.S. Virgin Islands, and certain Pacific Basin Islands. The target population covers all land ownership categories including federal land, although NRI data collection activities have historically concentrated on non-federal lands. The sample design used an area frame, which was deemed to better suit the NRI because the boundaries and definitions of sampling units may change.

The NRI employed a stratified two-stage area sampling design, where the PSUs are tracts of land, referred to as segments, while the SSUs are points within the segments. The segments are typically half-mile square parcels of land equivalent to 160-acre quarter-sections in the Public Land Survey System (PLLS). Most periodic inventories sampled approximately 300 000 segments and about 800 000 points, which is around 2 to 6 percent of total land area. It is only in 1987 that the sample was reduced to one-third, because of budget cuts. The segments are chosen within each strata by SRS without replacement, in which the strata are defined primarily based on geography, resource conditions and land ownership patterns.

Starting in 2000, the periodic NRIs were redesigned to accommodate the annualization of the inventory. Many of the design features of the previous NRI were retained in the new design. However, under the annualized inventory, a supplemented panel design was used. In such a design, a fraction of the segments, called the core panel, is observed every year to efficiently estimate the change parameters. The remaining fraction, called the supplemental panel, which is observed only part of the time, takes care of state estimates. The annualized inventory samples only about 40 000 segments as the core panel and 30 000 segments as supplemental panel every year since its implementation. Despite being high-maintenance in terms of budget, the annual inventory was embraced by the federal statistical system as it provides more timely information and better-quality annual operations.

The current NRI design, which was introduced in 2017, follows the annual inventories in terms of core and rotation structure. Data collection now includes interpretation of aerial photographs of sampled segments by three remote sensing laboratories and the use of tablets for navigation, point location and demarcation of field boundaries. The new design employs a hierarchical local pivotal method to obtain the sample. At present, planning for an optimal spatial-temporal design for model-based estimators and incorporating cost structure to the design remain to be worked on.

As for the data collection in 1987, NRI used remote sensing techniques to update the 1982 conditions for approximately 30 percent of the sample sites. The use of remote sensing increased during the 1990s. Starting in 2000, a special high-resolution imagery was acquired for each NRI sample site selected for that year's annual sample. Moreover, as part of the 2004 NRCS reorganization, the agency established remote sensing laboratories to take full advantage of modern geospatial technologies to facilitate efficient collection and processing of NRI data.

Annex 3

Methods and examples on sampling and estimation

Simple Random Sampling (SRS)

This is the basic sampling technique in which each element in the population has an equal probability of selection and each combination of elements has an equal probability of selection. To implement SRS, each element in the sampling frame is assigned a random number. The sampling frame is then sorted according to the random numbers. The first n elements in the sorted file are then taken as the sample.

Suppose that y_i is the value of the characteristic of interest for the i^{th} element. Then, the sample mean $\bar{y} = \frac{1}{n}\sum_{i \in S} y_i$ is an unbiased estimator for the population mean $\bar{Y}$.

The variance of the sample mean is $\text{Var}\left(\bar{y}\right) = \left(1 - \frac{n}{N}\right)\frac{S^2}{n}$,

where S^2 is the population variance. It is usually unknown; therefore, the variance of the sample mean is estimated by $\text{var}\left(\bar{y}\right) = \left(1 - \frac{n}{N}\right)\frac{s^2}{n}$. ·

Example

To estimate the total number of beetles in a large agricultural field, the field was subdivided into 100 equally sized units. An SRS of eight units was taken and all the beetles in these eight units were counted. The observed samples at the eight fields are: 234, 256, 128, 245, 211, 240, 202 and 267.

1	11	21	31	41	51	61	71	81	91
2	12	22	32	42	52	62	72	82	92
3	13	23	33	43	53	63	73	83	93
4	14	24	34	44	54	64	74	84	94
5	15	25	35	45	55	65	75	85	95
6	16	26	36	46	56	66	76	86	96
7	17	27	37	47	57	67	77	87	97
8	18	28	38	48	58	68	78	88	98
9	19	29	39	49	59	69	79	89	90
10	20	30	40	50	60	70	80	90	100

Estimate the total number of beetles in the large agricultural field and the corresponding standard error.

The sample mean number of beetles is $\bar{y} = 222.875$ and the sample variance is $s^2 = 1932.696$. Hence, the variance of the mean number of beetles is:

$$\text{var}(\bar{y}) = \left(1 - \frac{8}{100}\right)\frac{s^2}{8} = 222.26 \text{ and the standard error is } \text{se}(\bar{y}) = \sqrt{222.26} = 14.908$$

The estimated number of beetles for the large field is $N\bar{y} = 100 \times 222.875 = 22287.5$

The corresponding standard error is $\text{se}(N\bar{y}) = N\text{se}(\bar{y}) = 100 \times 14.908 = 1490.8$

Example: sample size determination

What sample size is needed to estimate the total population of beetles such that the estimate is within 1 000 of the true population total, at 95 percent CI?

$$P\left(\left|N\bar{y} - N\bar{Y}\right| < 1000\right) = 1 - 0.05 \Rightarrow P\left(100\left|\bar{y} - \bar{Y}\right| < 1000\right) = 1 - 0.05$$

$$= P\left(\left|\bar{y} - \bar{Y}\right| < 10\right) = 0.95$$

$$n_0 = \left(\frac{Z_{\alpha/2}S}{e}\right)^2 = \left(\frac{1.96 \times \sqrt{1932.696}}{10}\right)^2 = 74.246$$

$$n_{SRS} = \frac{n_0}{1 + \frac{n_0}{N}} = \frac{74246}{1 + \frac{74.246}{100}} = 42.61$$

Stratified sampling

When it is necessary to ensure that certain characteristics of the population are well represented in the sample (for example, irrigated and non-irrigated farms), each population unit may be classified according to the categories of a characteristic of interest and sample from each category. This approach, which is usually called stratification, ensures that the categories are represented in the sample.

Stratification improves the precision of estimates compared to SRS if the groups are different from each other and the elements within groups are homogeneous. The population can be stratified by any characteristic of interest that is available for all units in the sampling frame (such as geographical location – province, district, farm size, irrigated/upland). Because the auxiliary variable or the characteristic of interest used for grouping the population units must be observed or must be known for all population units, stratification is deemed costly compared to SRS. Therefore, there is a tradeoff between the cost of doing the stratification and the smaller sample size needed for the same error. Different sampling techniques may be used to draw the sample from the different strata or groups. When the population is skewed, the small groups may be oversampled to improve inter-group comparisons. Examples of skewed population groups are female-headed versus Male-headed farming households; and large farms versus smallholder farms

Supposing that there are H groups or strata and that a sample of size n_h is taken from N_h population units in stratum h. Then the population mean $\bar{Y}$ will be estimated by $\bar{y}_{st} = \sum_{h=1}^{H} W_h \bar{y}_h$, $W_h = N_h/N$ is the stratum h population proportion, $\bar{y}_h$ is stratum sample mean and is unbiased for population stratum mean $\bar{Y}_h$. The variance of $\bar{y}_{st}$ will be

$$\text{Var}(\bar{y}_{st}) = \sum_{h=1}^{H} W_h^2 \text{Var}(\bar{y}_h)$$

Example: stratified random sampling

The comprehensive reporting system of a small country for the production per hectare of its staple crop, palay, is summarized below by type of farms, irrigated and upland.

Farms	Population	Stratum 1 (Irrigated)	Stratum 2 (Upland)
Size	N = 1,000,000	N_1 = 200,000	N_2 = 800,000
Variance	S^2 = 1,800,000	S_1^2 = 4,000,000	S_2^2 = 1,000,000
Mean	$\bar{Y}$ = 1,400	$\bar{Y}_1$ = 3,000	$\bar{Y}_2$ = 3,000

Suppose that 1 200 farms were selected using proportional allocation and that SRS will be employed to select the farms. Estimate the possible variance of the production per hectare of palay.

$$n_1 = 1200 \times 200,000/1,000,000 = 240$$
$$n_2 = 1200 \times 800,000/1,000,000 = 960$$
$$W_1 = N_1/N = 200,000/1,000,000 = 0.2$$
$$W_2 = N_2/N = 800,000/1,000,000 = 0.8$$

$$\begin{aligned}\text{Var}(\bar{y}_{st}) &= W_1^2\left(1-\frac{240}{200000}\right)\frac{S_1^2}{240} + W_2^2\left(1-\frac{960}{800000}\right)\frac{S_2^2}{960} \\ &= (0.2)^2 \times 4000000/240 + (0.8)^2\, 1000000/960 \\ &= 666.7 + 666.7 = 1333\end{aligned}$$

Suppose that instead of stratified random sampling, SRS is used to select the 1 200 farms. What is the variance of the production per hectare estimate?

$$\text{Var}\left(\bar{y}_{SRS}\right) = \left(1-f\right)\frac{S^2}{n} = \left(1-\frac{1200}{1000000}\right)\frac{1800000}{1200} = 1500$$

Which is more precise: stratified random sampling or SRS?

$$D^2\left(\bar{y}_{st}\right) = \frac{\text{Var}\left(\bar{y}_{st}\right)}{\text{Var}\left(\bar{y}_{SRS}\right)} = \frac{1333}{1500} = 0.889$$

As the design effect of a stratified random sample is 0.889<1, then the stratified random sample is more precise. To achieve the same variability as that of stratified random sample, an SRS needs at least 1 200/0.889 = 1 350 observations.

Systematic sampling

Another sampling technique that is applicable in selecting elements and clusters is systematic sampling, in which the population units are arranged or sorted according to an auxiliary variable that is available in the sampling frame. Elements are the most basic population units, while clusters consist of elements that are grouped in proximity to one another or in a certain structure, such as farms (elements) in a village.

A sampling interval, *F*, is determined such that *F = N/n = 1 / f*, where *N* and n are the population and sample sizes, respectively. A random number is then drawn between *1* and *F*, *RN*. The sample will consist of the population units that have the following order numbers: *RN, RN + F, RN + 2F, RN + (n - 1)F.* If *RN + (n - 1)F* is larger than the population size, then the population unit corresponding to the order number *(RN + (n - 1)F - N)* will be included in the sample. This approach is often referred to as circular systematic sampling. Another approach that could be implemented when F is not an integer is to select the sample using the rounded down fractional interval, *RN* + kF, k = 0,1,2, n -1.

As there is only one random start that is used in selecting the sample, technically, the variance of any estimate from the sample cannot be computed. The variance of an estimate is usually estimated using the variance under SRS or using paired selection or successive difference models.

Paired selections — It is a reasonable assumption that each successive pair of selections was drawn at random, two from each implicit "stratum" or "zone". The number of zones will be *n/2*. The pairs of selections compared are determined by the order of selections.

If n is even, $\text{var}\left(\bar{y}_{sy}\right) = \frac{(1-f)}{n^2}\sum_{k=1}^{n/2}\left(y_{ka}-y_{kb}\right)^2$. If n is odd, select one of the sampled element at random and use it twice making $(n+1)/2 = m'$ pairs. Then $\text{var}\left(\bar{y}_{sy}\right) = \frac{(1-f)}{n(n+1)}\sum_{k=1}^{m'}\left(y_{ka}-y_{kb}\right)^2$

Successive difference model — This modifies the paired selection model and uses all (n-1) successive differences, that is, (1-2), (2-3), (3-4)…({n-1},n). The precision of the competed variance is somewhat increased. The degrees of freedom are larger than n/2 but lower than (n-1).

$$\text{var}\left(\bar{y}_{sy}\right) = \frac{(1-f)}{2n(n-1)}\sum_{g=2}^{n}\left(y_g-y_{g-1}\right)^2$$

Example: systematic sampling

There are 40 farms in the population and a sample of eight farms was selected using systematic SRS. The sampling interval $k = N/n = 40/8 = 5$. A random number between 1 and the sampling interval k was drawn, for example 5. The following are the answers to the question of how many parcels of land the selected farms constitute: 1, 2, 1, 2, 3, 4, 5, 3.

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16
17	18	19	20
21	22	23	24
25	26	27	28
29	30	31	32
33	34	35	36
37	38	39	40

$$\bar{y}_{sy} = 2.625 \quad f = 1/5 \quad (1-f) = 0.8$$

To estimate the variance of the sample mean, two models may be explored:

Paired selection model: $$\operatorname{var}(\bar{y}_{sy}) = \frac{(1-f)}{n^2}\sum_{k=1}^{n/2}(y_{ka} - y_{kb})^2$$

$$\operatorname{var}(\bar{y}_{sy}) = \frac{0.8}{8^2}\left[(1-2)^2 + (1-2)^2 + (3-4)^2 + (5-3)^2\right] = 0.0875$$

Successive difference model: $$\operatorname{var}(\bar{y}_{sy}) = \frac{(1-f)}{2n(n-1)}\sum_{g=2}^{n}(y_g - y_{g-1})^2$$

$$\operatorname{var}(\bar{y}_{sy}) = \frac{0.8}{2\times8\times7}\begin{bmatrix}(1-2)^2 + (2-1)^2 + (1-2)^2 + (2-3)^2 \\ +(3-4)^2 + (4-5)^2 + (5-3)^2\end{bmatrix} = 0.07143$$

Cluster sampling

In developing countries, small holdings are prevalent. Even if a comprehensive list of these holdings is available, directly selecting them using SRS or systematic sampling will not be cost-effective as it is quite likely that the resulting sample holdings may be scattered widely and will require higher travel costs for enumerators and supervisors. It will also be difficult to achieve a more balanced field operation because the work load of enumerators and supervisors assigned to different areas may not be uniform. A more viable approach would be to first select areas and then choose holdings in the selected areas. This method is known as cluster sampling, wherein elements, in this case holdings, are physically grouped into clusters that could be villages or census EAs. Field operations will then be limited to the selected clusters in which selected farms are relatively near to one another and thus travel costs will be lower. However, because farms that are in a village are more likely to share socio-economic characteristics, enumerating all of them in the selected clusters is not advisable as additional observations from the same clusters will not increase the precision of estimates. As an example, consider the design effect $D^2(\bar{y})$ of the estimator $\bar{y}$, which is the ratio of the variance of $\bar{y}$ in cluster sampling, $Var_{cl}(\bar{y})$ and the corresponding variance under SRS, $Var_{SRS}(\bar{y})$. The estimator in cluster sampling is more efficient if the design effect is less than or equal to 1. In cluster sampling, the design effect is

$$D^2(\bar{y}) = \frac{\text{Var}_{\text{cl}}(\bar{y})}{\text{Var}_{\text{SRS}}(\bar{y})} = 1 + (b-1)\rho,$$

where b is the number of elements that is sampled from a cluster and ρ is the intraclass correlation, which by definition is the Pearson's correlation coefficient of all pairs of elements within a cluster (Kish, 1965). The design effect will only be equal to 1 if $b = 1$ and will be less than 1 if ρ is negative. With $b = 1$, cluster sampling will be the same as *SRS*. A negative intraclass correlation implies that the elements in the clusters are negatively related, which is not a realistic assumption.

A cost-effective approach that will also render reliable estimates is to sample clusters and then sample elements within the clusters, making the sampling design two-stage. In a crop production survey, the first stage or PSUs are the villages or census EAs and the elements, or SUs, are the farms or holdings. To increase precision, PSUs can be stratified. If stratified cluster sampling is implemented, then a sample allocation method as described above can be applied.

In general, the PSUs are well defined with clear and stable boundaries. Information must be available on the estimated sizes of the PSUs (to be used in probability proportional to size sampling) and on the characteristics of the PSUs (to be used for stratification). There should be a sufficient number of PSUs from which to select the sample.

If the same set of sampled PSUs will be used for several surveys (a master sample), the PSUs need to be large enough in terms of holdings to support all the agricultural surveys for which the sample can be used. The inclusion of a holding in more than one survey may cause a heavy response burden for that agricultural operator and may give rise to a decline in willingness to cooperate on the second occasion. Apart from survey designs that call for repeat interviews and observations at the same holding, different surveys should sample different holdings.

Example: cluster sampling (equal sized clusters)

An agricultural survey was conducted in district with 180 communes and 12 villages are located in each commune. An SRS of 20 communes was selected. In each sampled commune, all the 12 village chiefs were interviewed. One of the questions asked is whether the village has communal irrigation or not. The counts of the villages with communal irrigation are as follows: 2, 4, 2, 7, 9, 1, 0, 8, 1, 2, 3, 1, 4, 9, 1, 3, 4, 2, 8, 7

a. Estimate the proportion of villages with irrigation and its standard error.

From this sample: $p_c = \frac{\sum_{\alpha} y_{\alpha}}{n} = \frac{78}{20 \times 12}$, where $n = aB$

$p_c = 78/240 = 0.325 = 0.05252$ or 32.5%

32.5% of the villages have communal irrigation.

$$s_a^2 = \frac{1}{a-1}\sum (p_{\alpha} - p_c)^2 = 0.06206$$

$$\text{var}(p_c) = (1-f)s_a^2/a = (1-20/180)0.06206/20 = 0.002758$$

$$\text{se}(p_c) = \sqrt{0.002758} = 0.05252 = 5.3\%$$

b. Estimate the proportion of villages with irrigation and its standard error.

$$p_0 = 78/240 = 0.325$$

$$\text{var}(p_{SRS}) = (1-f)\frac{p_0(1-p_0)}{n-1} = \left(1 - \frac{240}{2160}\right)\frac{0.325 \times (1-0.325)}{239} = 0.0008159$$

For the cluster sample $\text{var}(p_c) = 0.002758$. The design effect for p_c:

$$d^2(p_c) = \frac{\text{var}(p_c)}{\text{var}(p_{SRS})} = \frac{0.002758}{0.0008159} = 3.381$$

The effective sample size is $n_{eff} = 240/3.381 = 70.99 \approx 71$

c. The survey is to be repeated with an SRS of 40 communes and a SRS of six villages in each sampled commune, giving the same sample size of n = 240. For this new design, estimate the design effect for the proportion of villages with communal irrigation and also estimate the standard error of the proportion.

With $d^2(\bar{y}) = 3.381$ and $B = 12$

$$\hat{\rho} = \frac{d^2(\bar{y}) - 1}{B-1} = \frac{3.381-1}{11} = 0.21642$$

$$\Rightarrow d^2(\bar{y}) = 1 + (b-1)\hat{\rho} = 1 + 0.21642(5) = 2.08212$$

$$\text{var}(p_c) = 2.08212 \times \text{var}(p_0) = 0.00169$$

$$\text{se}(p_c) = \sqrt{0.00169} = 0.041216$$

Selecting PSUs and USUs

With a two-stage sampling design, the probability of an element or USU β in a given cluster or PSU α of being in the sample is $P(\alpha\beta) = P(\alpha)\,P(\beta|\alpha)$. In a domain d if b USUs will be drawn from a PSUs and there are A PSUs of varying sizes such that M_a is the measure of size for PSU α, assuming that both PSUs and USUs were selected using SRS, then

$$P(\alpha\beta) = P(\alpha)P(\beta|\alpha) = \frac{a}{A}\frac{b}{M_\alpha} = f_d$$

This implies that the selection probabilities will vary across clusters, as well as the base weights, which are the inverse of the selection probabilities. A large coefficient of variation of the base weights is indicative of a larger contribution of the base weights to the loss of precision of the estimate. Thus, maintaining a uniform selection probability within the domain is beneficial. To achieve this under the same condition of selecting PSUs and USUs using SRS, the number of USUs to be selected must not be fixed, as b should vary with M_a. This approach, however, may not render a balanced workload among data collectors and may result in a more complicated field operation.

PPS sampling

A better way to achieve EPSEM in a two-stage cluster sampling design is to use PPS to sample PSUs. To illustrate, in a domain d, to maintain a uniform sampling fraction $f_d = n_d / N_d$, where n_d is the number of total USUs to be sampled, which is $n_d = a \times b$ and N_d is the total number of sampling units which is $\sum_{\alpha=1}^{A} M_\alpha$,

$$P(\alpha\beta) = P(\alpha)P(\beta|\alpha) = P(\alpha)\frac{b}{M_\alpha} = f_d = \frac{n_d}{N_d} \Rightarrow P(\alpha) = \frac{n_d}{N_d}\frac{M_\alpha}{b} = \frac{aM_\alpha}{\sum_{\alpha=1}^{A} M_\alpha}$$

Example: PPS sampling

In a stratum of an agriculture survey, two villages and 50 farms from each selected village will be drawn. The villages are of varying sizes. Examine this stratum and identify issues that need to be resolved before sample selection.

Village	M_a (No. of farms)	Cumulative M_a
A	300	300
B	30	330
C	200	530
D	900	1 430
E	100	1 530
F	70	1 600

Issue: oversized village

Village *D* is bound to be selected and has a one-eighth chance of being selected twice. With $a = 2$ and $b = 50$

$$f = \frac{2M_\alpha}{1600}\frac{50}{M_\alpha} = \frac{1}{16}$$

The probability of selecting village D, however, is:

$$P(\text{village D}) = \frac{2M_\alpha}{1600} = \frac{2\times 900}{1600} = 1\frac{1}{8} > 1$$

Solution: treat village D as a self-representing PSU. Thus, the stratum is collapsed into two strata, one with Village D only while the second stratum will contain the rest of the villages. Refer to the table below:

Village	M_a	Cumulative M_a
A	300	300
B	30	330
C	200	530
E	100	1 530
F	70	1 600
D	900	1 430

To maintain EPSEM for the two strata, adjust *b* such that $\frac{1}{16} = \frac{b}{900} \Rightarrow b = \frac{900}{16} = 56.25.$ Hence, for Village D, the number of farms to be selected will be 56 instead of 50.

For the second stratum with the remaining villages, the selection equation will be: $\frac{1}{16} = \frac{M_\alpha}{700} \frac{b}{M_\alpha} \Rightarrow b = \frac{700}{16} = 43.75.$

Thus, for the second stratum, 44 farms will be selected from the sample village.

Issue: undersized PSU

In the second stratum, village B is undersized. If village B is selected, only 30 farms can be included in the sample, instead of 44. In this case, village B should be collapsed with an adjacent village, for example village A. Therefore, if village A+ B is selected, the list of farms from which to perform the selection should include all the farms in both village A and B.

Determining the number of PSUs and USUs to be selected

In a two-stage sampling for a given domain, when the sample size n_d has been determined, the number of USUs to be selected per cluster will be determined next; then, the number of clusters to be selected can be derived. If the unit cost of collecting data from a holding, c, and the unit cost of surveying a PSU, C_a and an estimate of the intraclass correlation, $\hat{\rho}$ can be computed from previous similar surveys, then the optimum number of USUs to be selected that will render a minimum variance of the sample mean subject to a fixed cost $C = aC_a + nc = aC_a + a \times b \times c$ is

$$b_{opt} = \sqrt{\frac{C_a}{c} \frac{(1-\hat{\rho})}{\hat{\rho}}}\ .$$

In a simplified context, in domain d, the sampling fraction f_d would be computed as n_d/N_d, where n_d is the required sample size and N_d is the population size. In practice, however, the computation of f_d is more complicated because allowance needs to be made for nonresponse. The allowance for unit (holding or farm) nonresponse is made by increasing the size of the selected sample so that it will yield a responding sample of the specified sample size, n_d.

The selected, or adjusted, sample size per domain can be calculated using a domain-level nonresponse rate computed from previous surveys, if any. The adjusted sample size in a domain denoted by n^*_d can be computed as $n^*_d = n_d / \bar{m}_d$, where n_d is the desired responding sample size for region d and $\bar{m}_d$ is the response rate for that domain.

Example: finding an optimum number of PSUs and USUs

A farming household survey on the willingness to pay for irrigation will be conducted in remote villages that do not have access to communal irrigation. The intraclass correlation from previous surveys was estimated to be 0.10 and the projected per cluster surveying cost is approximately 3 050 Philippine Pesos (PHP), while the cost of surveying a farming household is PHP 350. Find the optimum number of households and the optimum clusters to be sampled if the total survey operations budget will be 1 million Php.

$$b_{opt} = \sqrt{\frac{C_a}{c}\frac{(1-\rho)}{\rho}} = \sqrt{\frac{3050}{350}\frac{(1-0.1)}{0.1}} = \sqrt{8.714\times 9} = 8.856 \approx 9$$

$$\text{Since } C = aC_a + nc \Rightarrow C = aC_a + abc \Rightarrow a = c/(C_a + bc)$$

$$a = \frac{1{,}000{,}000}{3050 + 350\times 9} = 161.2903 \approx 161$$

Replicated or interpenetrating samples

Deriving estimates and their corresponding variances can become complicated for surveys with more than one stage of selection. To simplify variance estimation, Mahalanobis introduced replicated sampling, which employs c independent small samples instead of a big sample (Kish, 1965). Unbiased estimates of the population mean, $\bar{Y}$, can be derived from each of these independent probability samples, such as $\bar{y}_\gamma$, such that the composite estimate $\hat{\bar{y}} = \sum_{\gamma=1}^{c} \bar{y}_\gamma \Big/ c$ is also unbiased. The corresponding variance of $\hat{\bar{y}}$ will be

$$\text{var}(\hat{\bar{y}}) = \frac{1}{c(c-1)}\sum_{\gamma=1}^{c}(\bar{y}_\gamma - \hat{\bar{y}})^2$$

which is relatively straightforward.

Use of a replicated sample also addresses the problem of variance estimation for a systematic sample with only one random start. Instead of a large sampling interval F, c samples can be drawn such that each sample will have a sampling interval $k = F / c$.

For surveys that are conducted regularly or periodically, rotation of PSUs or USUs can be easily introduced in a replicated sample. Also, when the budget for a full sample (all the replicates are included) is not available, the survey can be conducted using a subset of the replicated sample, without complicated adjustments to the sampling design.

Simple replicated sample: example

The master sample of a province consists of four replicates, each selected with a sampling interval of 60. For each of the replicates, the total number of workers and farmers (using the FAO definition) were reported in the following table:

Replicate	No. of workers	No. of farmers
1	4 995	355
2	4 880	375
3	4 500	330
4	4 900	300

1. Estimate the total number of farmers and its standard error.
2. Estimate the proportion of farmers and its standard error.
3. If the total number of workers is actually 280 000, estimate the total number of farmers. Compare this from the estimate from (1).

Replicate, i	x_i	y_i	$\hat{p}_i = y_i / x_i$	$\hat{y}_i = ky_i$
1	4 995	355	0.07107	21 300
2	4 880	375	0.07684	22 500
3	4 500	330	0.07333	19 800
4	4 900	300	0.06122	18 000

Sampling interval of each replicate is $k = 60. \Rightarrow cF = 60 \Rightarrow F = 15, f = 1/15$

Hence, for each replicate, the estimate of the total $\hat{y}_i = ky_i = 15y_i$

$$\hat{Y} = \frac{1}{c}\sum_{i=1}^{c}\hat{y}_i = \frac{1}{4}\sum_{i=1}^{4}60y_i = \frac{81600}{4} = 20,400$$

$$\text{var}(\hat{Y}) = \frac{(1-f)}{c(c-1)}\sum_{i=1}^{4}(\hat{y}_i - \hat{Y})^2 = \frac{(1-1/15)}{4(4-1)} \times 1140000 = 882000$$

$$\text{se}(\hat{Y}) = \sqrt{882000} = 939.149$$

Compute $\hat{p}_i = y_i / x_i$ for each replicate i and then $\hat{p} = \sum \hat{p}_i / c$.

$$\hat{p} = \frac{1}{c}\sum_{i=1}^{c}\hat{p}_i = \frac{1}{4}0.28247 = 0.070618$$

$$\text{var}(\hat{p}) = \frac{(1-f)}{c(c-1)}\sum_{i=1}^{c}(\hat{p}_i - \hat{p})^2 = \frac{(1-1/15)}{4(4-1)}0.000135 = 1.04675 \times 10^{-5}$$

$$\text{se}(\hat{p}) = 0.00323$$

$$\hat{Y} = X \times \hat{p} = 280000 \times 0.070618 = 19,773.121$$

$$\text{se}(\hat{Y}) = X \times \text{se}(\hat{p}) = 280000 \times 0.00323 = 905.900$$

Example: weighting adjustments for nonresponse

A stratified multi-stage sample of 950 farm holdings is selected with farms in the north (N) being sampled at the rate of 1/100 and those in the south (S) at the rate of 1/200.

The response rate in the lowland areas is lower than in the upland areas.

Let n_h be the issued sample, n'_h be the achieved sample and r'_h be the number of farms planted to rice. The weight will be the product of the base weight and adjustment for nonresponse. For each stratum h $w_h = w_{h1} \times w_{h2}$ $\quad w_{h2} = n_h / n'_h$

Adjustment Cells	n_h	n'_h	r'_h	w_{h1}	$w_{h2.1}$	w_h	$w_h n'_h$	$w_h r'_h$
NL	100	70	60	100	1.43	143	10 010	8 580
NU	300	270	140	100	1.11	111	29 970	15 540
SL	150	120	110	200	1.25	250	30 000	27 500
SU	400	380	210	200	1.05	210	79 800	44 100
	950	840					149 780	95 720

The estimated proportion of farm holdings planted to rice is:

$$r = \frac{\sum w_h r_h}{\sum w_h n'_h} = \frac{95,720}{149,780} = 0.639.$$

In the previous example, suppose that the number of farm holdings is known to be 44 080 for N and 107 920 for S. The weighted sample numbers are 39 980 for N and 109 800 for S. To correct the weighted sample distribution to make it conform to the population distribution, the weight of sampled farms needs to be changed by the factors:

$$N: w_{h3} = \frac{44{,}080}{39{,}980} = 1.103 \quad S: w_{h3} = \frac{107{,}920}{109{,}800} = 0.983$$

The final weight is then $w_h = w_h \times w_{h3}$

Adjustment cells	n'_h	r'_h	w_h	w_h	$\tilde{w}_h n'_h$	$\tilde{w}_h r'_h$
NL	70	60	143	157.7	11 039	9 642
NU	270	140	111	122.4	33 048	17 136
SL	120	110	250	245.7	29 484	27 027
SU	380	210	210	206.4	78 432	43 344
	840				152 003	96 969

With the final weights $\tilde{w}_h$, the weighted sample number of farm holdings in *N* and *S* are, as required, 44 080 and 107 920 respectively.

Proportion of farms planted to rice: $\frac{96{,}969}{152{,}004}$=0.638.

Example: ratio estimation

An experienced farmer makes an eye estimate of the weight of oranges, x_i on each tree in an orchard of $N=200$ trees. He finds a total weight of $X=11\ 600\ kg$ and would like to use this estimate for selling his produce in bulk. However, the buyer would like to verify this estimate. He bought the fruits of ten rambutan trees selected randomly. The fruits in the ten trees were picked and weighed to verify the accuracy of the farmer's eye estimate. Below is the summary of the results.

Tree No.	*i*	1	2	3	4	5	6	7	8	9	10	Total
Actual weight	y_i	61	42	50	58	67	45	39	57	71	53	543
Eye Estimate	x_i	59	47	52	60	67	48	44	58	76	58	569

On the basis of the summary above, help the buyer determine the precision of the ratio of the actual weight to the owner's eye estimate, estimate the total weight of oranges that will be harvested from the 200 trees and then estimate the corresponding standard error.

Based on the sample, the ratio of actual weight to the owner's estimate is: $r = \frac{\bar{y}}{\bar{x}} = \frac{y}{x} = \frac{543}{569} = 0.9543$

Therefore, based on the sample, the total weight of oranges that will be harvested will be:
$\hat{y}_r = X\ x\ r = 11\ 600\ x\ 0.9543 = 11\ 069.95\ kg$. This is a better estimate than the owner's eye estimate because it is based on the actual weighing of the produce of ten randomly selected trees.

Because the sampled trees were randomly selected, the formulae of SRS can be used to compute for the variances of the means of actual weights, $\bar{y}$, eye estimate, $\bar{x}$ and covariance of the actual weight and eye estimate.

$$s_y^2 = 110.9 \quad s_x^2 = 94.54 \quad s_{xy} = 99.77$$

$$\text{var}(\bar{y}) = \left(1 - \frac{n}{N}\right)\frac{s_y^2}{n} = \left(1 - \frac{10}{200}\right)\frac{110.9}{10} = 10.4$$

$$\text{var}(\bar{x}) = \left(1 - \frac{n}{N}\right)\frac{s_x^2}{n} = \left(1 - \frac{10}{200}\right)\frac{94.54}{10} = 8.982$$

$$\text{cov}(\bar{y},\bar{x}) = \left(1 - \frac{n}{N}\right)\frac{s_{xy}}{n} = \left(1 - \frac{10}{200}\right)\frac{99.77}{10} = 9.47$$

From these computations, the variance of the ratio mean and, consequently, the estimate of the total harvest can be derived as follows:

$$\begin{aligned}\text{var}(r) &\approx \frac{1}{n\bar{X}^2}\left\{\text{var}(\bar{y}) + r^2\,\text{var}(\bar{x}) - 2r\,\text{cov}(\bar{y},\bar{x})\right\}\\ &= \frac{1}{10}\left(\frac{200}{11600}\right)^2\left\{10.534 + 0.9543^2 \times 8.982 - 2 \times 0.9543 \times 9.47\right\}\\ &= 0.00019\end{aligned}$$

$$\text{var}(\hat{y}_r) = X^2\,\text{var}(r) = 11600^2 \times 0.00019 = 2509.43$$

$$\text{se}(\hat{y}_r) = \sqrt{2509.43} = 159.72$$

Also, evaluate the adequacy of the approximation of the standard error.

The approximation of the standard error above is adequate if *($\bar{x}$) = cv (x)* is less than 0.1 and it is tolerable if less than 0.2.

$$\text{cv}(\bar{x}) = \frac{\text{se}(\bar{x})}{\bar{x}} = \frac{\sqrt{8.982}}{56.9} = 0.0527$$

Thus, the approximation of the standard error is adequate.

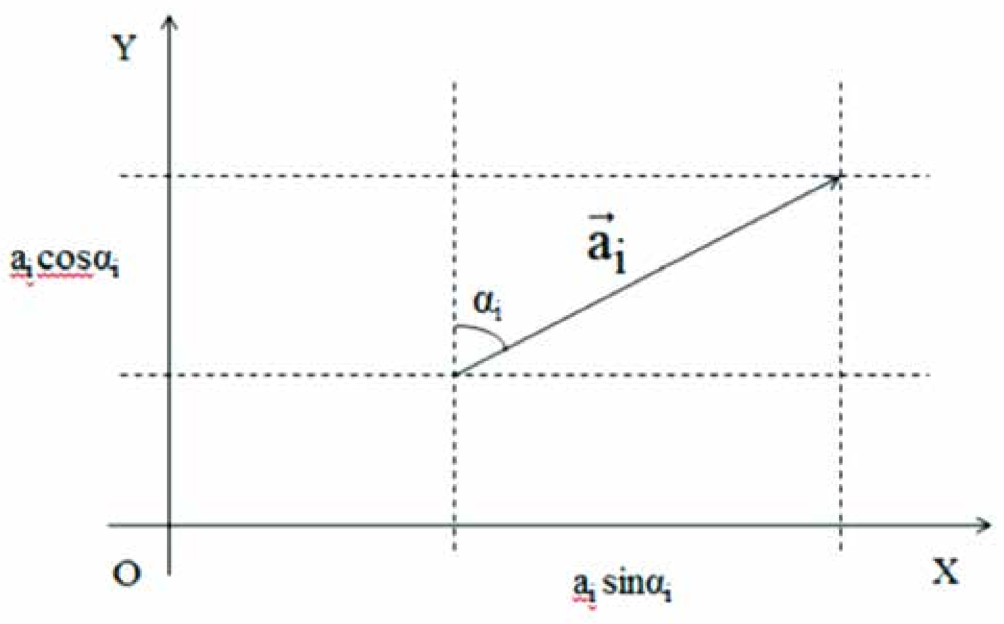
Annex 4

The polygon method

The procedure as given in FAO (1982) is provided here.

Let a polygon with n sides be defined as a_i, α_i, $i = 1,2\ n$, where a_i is the length of the side i and α_i is the angle formed by this side with north, measured in a clockwise direction. Let a vector a^I_i represent the side i in a two-dimensional space XOY in which the Y-axis coincides with the north. Thus, the horizontal and vertical projections of the vector a^I_i (Figure A4.1) are $a_i \sin \alpha_i$ and $a_i \cos \alpha_i$, respectively.

FIGURE A4.1 HORIZONTAL AND VERTICAL PROJECTIONS OF A VECTOR



Source: FAO, 1982.

Define vectors

$$\vec{R}_i = \sum_{j=1}^{i} a_j, \quad \forall i=1,2,...,n$$

Their horizontal and vertical projections will be

$$X_i = \sum_{j=1}^{i} a_j \operatorname{Sin} a_j \qquad \text{(A4.1)}$$

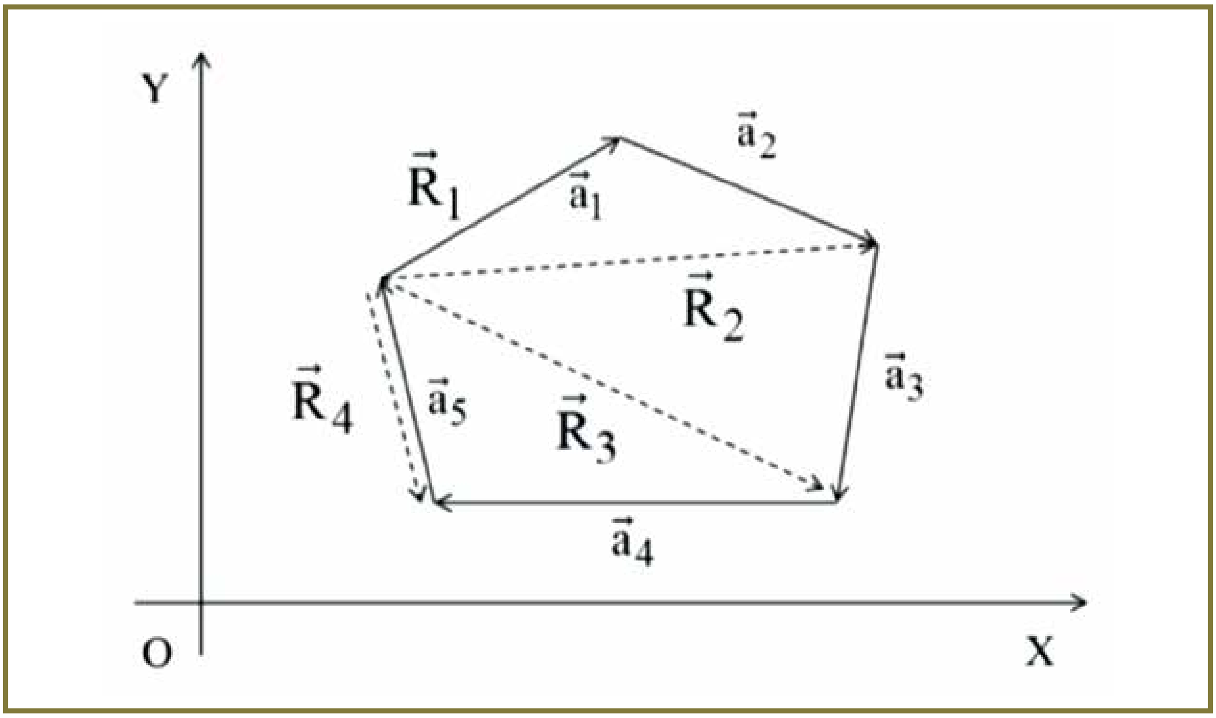
$$Y_i = \sum_{j=1}^{i} a_j \operatorname{Cos} a_j \quad \text{, respectively.} \qquad \text{(A4.2)}$$

If the polygon is closed, then $\vec{R}_n = 0$.

The area of a triangle formed by two vectors, which start from the same point, can be calculated as a function of their horizontal and vertical projections. Thus, the area of the triangle between vectors $\vec{R}_1$ and $\vec{R}_2$ (Figure A4.2) is given by

$$A_1 = \frac{1}{2}(X_2 Y_1 - X_1 Y_2) \qquad \text{(A4.3)}$$

FIGURE A4.2 AREA DETERMINATION



Source: FAO, 1982.

It should be noted that this area will have a positive value if the vector $\vec{R}_1$ precedes the vector $\vec{R}_2$ while looking clockwise; otherwise, it will have a negative value. The area of the whole polygon, calculated as a sum of the areas of triangles, each formed by the two consecutive vectors $\vec{R}_i$ is

$$A = \frac{1}{2}\sum_{i=1}^{n-2}(X_{i+1}Y_i - X_iY_{i+2}) \quad \text{(A4.4)}$$

where X_i and Y_i are given by equations (A4.2) and (A4.3).

Closure error and corrected area of a polygon

In practice, the polygon defined by the data collected in the field will never close. In this case, $\vec{R}_n \neq 0$.

The length of the vector $\vec{R}_n$ is given by $R_n = \sqrt{X_n^2 + Y_n^2}$

which can be used as a measure of error. However, a common practice is to express the closure error as a percentage of the perimeter of the polygon:

$$C = \frac{R_n}{\sum_{i=1}^{n} a_i} \times 100 \quad \text{(A4.5)}$$

The errors are considered acceptable if the closing error is less than 2 percent. There are different methods of closing the polygon (FAO, 1982), including:

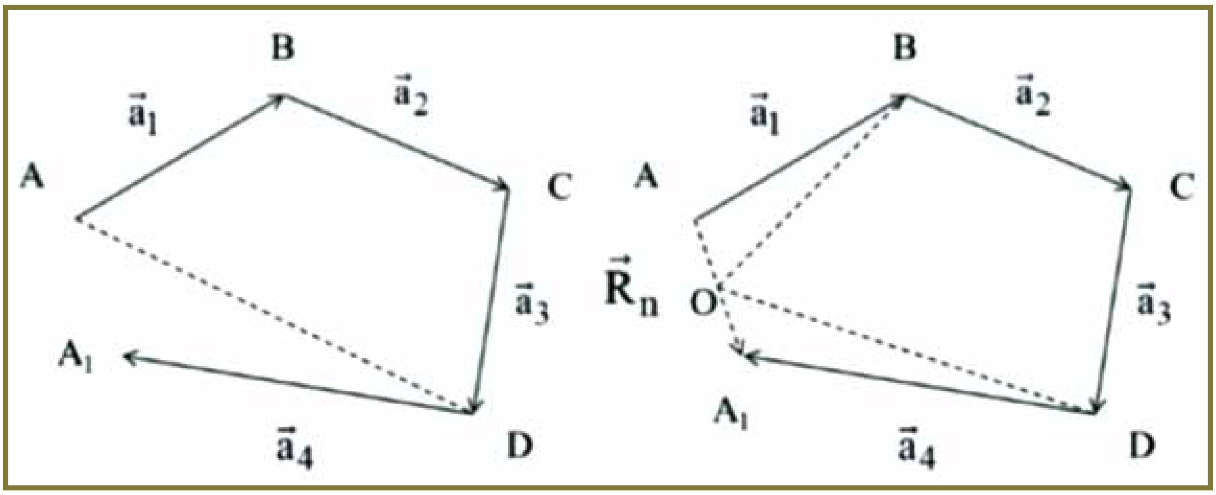
A. closure by connecting the last but one point with the starting point;

B. closure from the midpoint;

C. closure by shifting all vertices on an equal basis; and

D. closure by shifting all vertices on a proportionate basis.

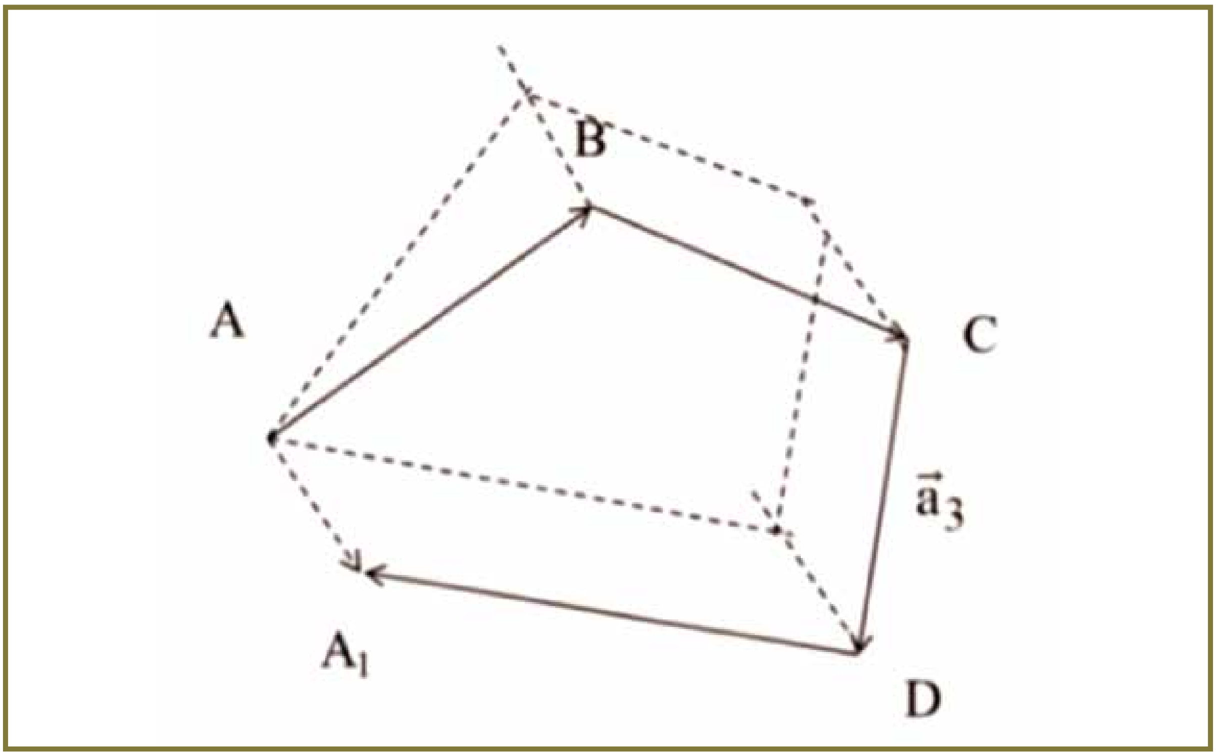
FIGURE A4.3 METHODS A, B AND C



Source: FAO, 1982.

An advantage of the first three methods is that there is no need to keep in the memory all input data until the end of the calculation. In these three methods, each pair of input data can be elaborated when they are entered, and required sums can be aggregated. The moment the last pair of data is entered, the corrected area and closure error can be evaluated. On the other hand, in the fourth method, there is a need to keep in the memory all input data for one polygon until the calculations are completed. It is important to note that the four methods give an unbiased estimate provided that measurement errors are absent.

FIGURE A4.4 METHOD D



Source: FAO, 1982.

Annex 5

Measuring distances

Some land surveying methods, such as rectangulation, triangulation and the P2/A method, described in chapter 4 require measurement of distances, whether the length or breadth of a rectangle, the base and the height of a triangle or a perimeter. Consequently, the methods for measuring distance form an important component of these methods. In developing countries, most farmers are unaware of the magnitude of the areas under the crops grown in their fields. Accordingly, field enumerators are employed to measure distances either by pacing or by using instruments.

1. Pacing

Pacing means walking at a normal gait and counting the number of steps required to cover a certain distance. The steps are then converted to standard units. To begin with, an average length of steps (usually 0.83 m) is used by the enumerators. As there may be individual differences in steps, the step of each enumerator needs to be calibrated by pacing a well-known distance and using that as the conversion factor.

It has been observed that the length of the pace of even the same enumerator changes with changes in the type of surface, such as sandy soil, clay and uneven ground. The length of the pace also varies with the enumerator's state of health and level of fatigue. Therefore, it has become necessary to calibrate the step several times a day, which makes pacing cumbersome.

There is also a risk of miscounting the number of paces, especially when this number is large. To eliminate this risk, use a simple instrument, the pedometer, was proposed. A pedometer, which consists of a digital reader and a dial, measures the movements of the body; each step taken is registered and shown on the dial. Each pedometer needs to be tested before use in case some of them may be out of order.

For the above reasons, the pacing method for measuring the length of sides or diagonals of a field is not recommended and, in fact, has been discontinued in almost all countries. However, it can be used, even without calibration, when random points are to be selected within a field for the purpose of laying crop-cutting plots.

2. Measuring distances with instruments

There is a range of instruments that are used for measuring distances. A commonly used method for measuring distances involves standardized cords and a wooden pole to a graduated wheel. Below are descriptions of how these instruments are used.

Standardized cord and wooden pole. A cord of 50 m has been used for the allotment of 50 m × 50 m parcels of communal land to village members in many African countries. To estimate crop areas, the cord needs to be of a non-extensible material and care should be taken to avoid getting it wet, or its length would be altered.

For example, a wooden pole, the kassaba, of 3.55 m, has been used to measure the sides of fields in Egypt. A common source of error encountered in large fields is the miscounting of the number of kassabas.

Surveyor's chain. The classical method for measuring distances is with a surveyor's chain. The metallic chain is composed of straight links with circular ends connected by rings with a handle. Each link is 0.20 m long measured from the centre of one connecting ring to the centre of the next. The usual chain length is 20 m (100 links), although there are 10 m and 50 m chains. Similar chains graduated in yards and feet are also available (FAO ,1982).

Two men are required to measure a distance, for example AB, with a chain: one man holds one end of the chain at point A while the other stretches the chain on the ground along the direction AB and marks the point A1, corresponding to the end of the chain. Then, the first man moves to point A1 and the operation is repeated as many times as necessary. The distance AB is calculated as so many complete chains plus a number of links.

While the chain is a cheap and secure instrument, it is heavy and not easy to handle. If not handled carefully, the links tend to bend, which reduces its length and results in overestimating the distance over a long distance. There is also the risk of forgetting to count a chain length. When the ground is uneven and the chain is not placed on the ground but held a few cm above, a slight error may be introduced due to the catenary effect.

Tapes. A low-cost instrument for measuring distance is metallic or plasticized tape, which is a substitute for the surveyor's chain. Tapes are wound on a special reel and are graduated in m, dm and cm or in yards, feet and inches. They are available in different lengths, such as 20 m, 30 m, and 50 m, or 50 feet and 100 feet. The distances are measured in the same way as with chains. The tape is not liable to bend and the catenary effect is almost nonexistent. It is easier to handle and generally more accurate. However, the tape is likely to break easily. It may rust if not cleaned after use and the plasticized tapes may lose their markings of the graduations.

Although the above measuring instruments (chains and tapes) are not very costly, the operating expenses are high, as two persons are needed to take the measurements. The remuneration of two persons, even if one of them need not be a professional enumerator but simply a labourer, in the long run, is more expensive than the use of more costly instruments that can be managed by a single person. Such instruments are the topofil, the Trumeter or Smith wheel and the optical range finder.

Topofil. The topofil is a distance measuring device fitted with a non-recoverable, light but strong string and a counter that registers distances in dm, m and hectometers. The string runs out of the instrument as the enumerator walks the distance to be measured. It can be easily carried and used by the enumerator. The enumerator fastens the end of the string to a fixed point and sets the counter at zero; as the enumerator walks the distance, the string unrolls and the counter registers the length of the string unrolled; at the terminal point, the enumerator reads on the counter the length of string unrolled, which is then cut and discarded. Any distance not longer than the length of the string on the reels can be measured in one single operation (maximum length 5.480 m) using the topofil. The speed of measurement matches the gait of the enumerator; the enumerator can read the counter at any intermediate point and set back the counter to zero, and as the distance is recorded mechanically, there is no danger of miscalculating long distances.

The apparatus is costly, with a high running cost as one reel can be used to measure at most 20 fields of small dimensions (about 5 ha).The topofil case is heavy and therefore inconvenient for carrying over long distances, and the string sags slightly and may even rest on the ground or on the plants.

Graduated wheel/Trumeter/Smith wheel. The main elements of a trumeter or a Smith wheel are a graduated wheel, a handle to push it and a counter, which registers the number of revolutions of the wheel. The circumference of the wheel is either 1 m or 1 yard. At the starting point, the enumerator sets the counter at zero and pushes the wheel along the line until the length of the distance to be measured. The reading on the counter plus the length corresponding to an incomplete revolution gives the length of the distance under consideration.

The cost of a graduated wheel is not very high and there are no running expenses. The instrument is easy to use and, therefore, does not require skilled enumerators. The accuracy of the instrument is high on smooth dry land, and the mechanical recording of the number of revolutions eliminates the risk of mistakes in counting.

The instrument is not ideal for use in a number of areas, including those with rough ground, ploughed land and irrigated and humid land. It cannot be used for direct measurement of horizontal distances when the land is sloping and when the land surface is undulating, as the wheel measures the wavy curve and not the straight line.

Annex 6

Crop cutting experiments

In conducting CCEs, it is expected that each enumerator will possess the following instruments or materials before he or she visits the selected field:

- Compass/GPS
- Measuring tape (30 m or 50 m)
- Calculator
- Rope or string (at least 30 m long)
- Weighing balance or spring balance
- Pegs (at least four and should be long)
- Bags to store the harvested crops and for keeping other items such as notebooks and questionnaires
- Baskets or containers to carry the harvested crops and for other activities
- Random number tables
- Cutlass or sickle to cut the crops
- Questionnaires, notebooks, instruction manuals and other relevant documents

FIGURE A6.1 TOOLS AND MATERIALS NEEDED IN THE FIELD FOR CONDUCTING CCES

Source: Author elaboration, 2018.

Based on the practical application of the crop cutting method in many countries, the following could be described as a summary of steps involved in conducting CCEs:

1. Assume that crop yield and production for a given area will be collected using the crop cutting method. If the area of interest is large, such as a province or a district, choosing holdings in which the crop cutting method will be conducted can be done efficiently using two-stage sampling. The area of interest can be divided into PSUs. PSUs can be groups of segments or areas, or geographical area such as villages with measure of size (for example, the number of holdings). PSUs are usually selected using PPS and in the second stage, holdings are selected either using SRS or systematic sampling. PSUs can be stratified according to the crops of interest to ensure that all the crops that need to be measured will be represented in the sample.
2. After selecting the holding, the next step is to determine the parcels that will be selected. In many developing countries, smallholders' holdings constitute more than one parcel that may be in different locations. Usually, all parcels in a selected holding are measured.
3. Demarcate the experimental plot of a given size and shape within the field or farm using appropriate random techniques.

Several studies have been conducted since the inception of the crop cutting method to examine the relative efficiency of plot sizes and shapes for yield rates. However, none of the studies provides any concrete evidence to indicate whether a particular plot shape is more efficient than the others, except for the well-known statistical fact that as the plot size increases, the crop cutting method approaches the whole plot harvest method. Various plot shapes and sizes that have been experimented in many countries, include a square plot of size 5 m x 5 m or 10 m x 10 m, a rectangular plot of size 10 m x 5 m, an equilateral triangle plot having sides of 10 m and a circular plot of radius 1.7145 m. The table below gives a summary of plot sizes and shapes for some food and non-food crops which were initially experimented in some states in India (Sud et al., 2015).

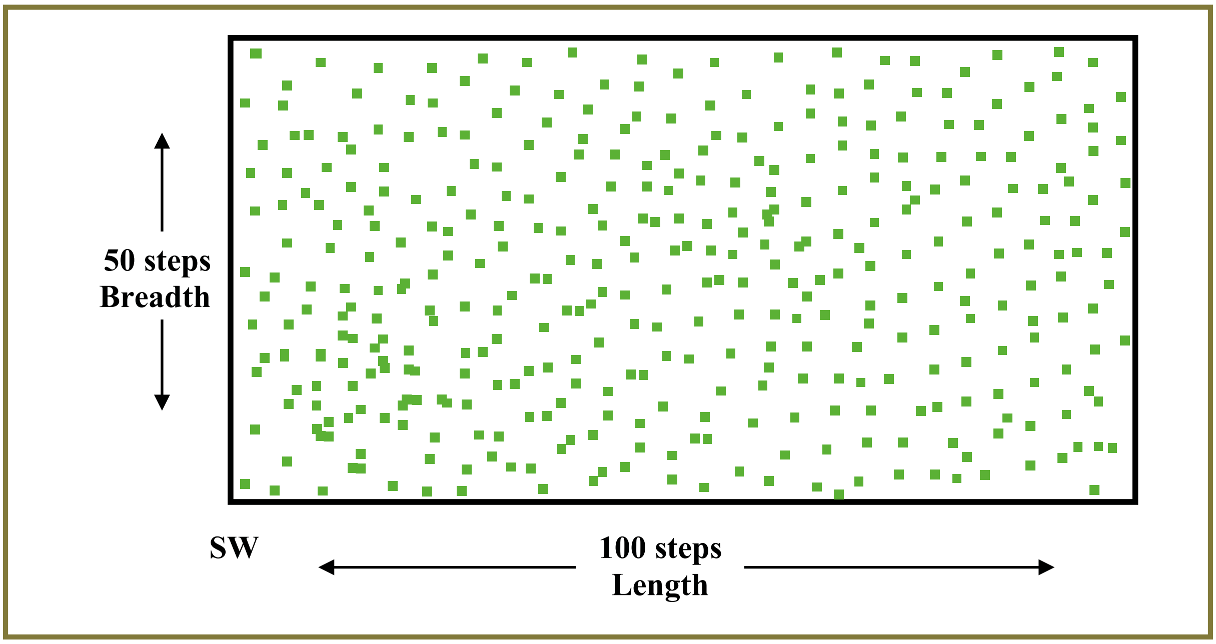
TABLE A6.2. PLOT SIZE AND SHAPE FOR SOME FOOD AND NON-FOOD CROPS.

Name of the crop	Shape	Length (m)	Breadth (m)	Diagonal (m)
Paddy, wheat, sorghum, pearl millet, maize, groundnut, tobacco, sugarcane, green gram, chilly, horse gram, black gram, chickpea, sunflower	Square	5	5	7.07
Redgram, sesamum, caster, cotton	Square	10	10	14.14

Source: Sud et al., 2015

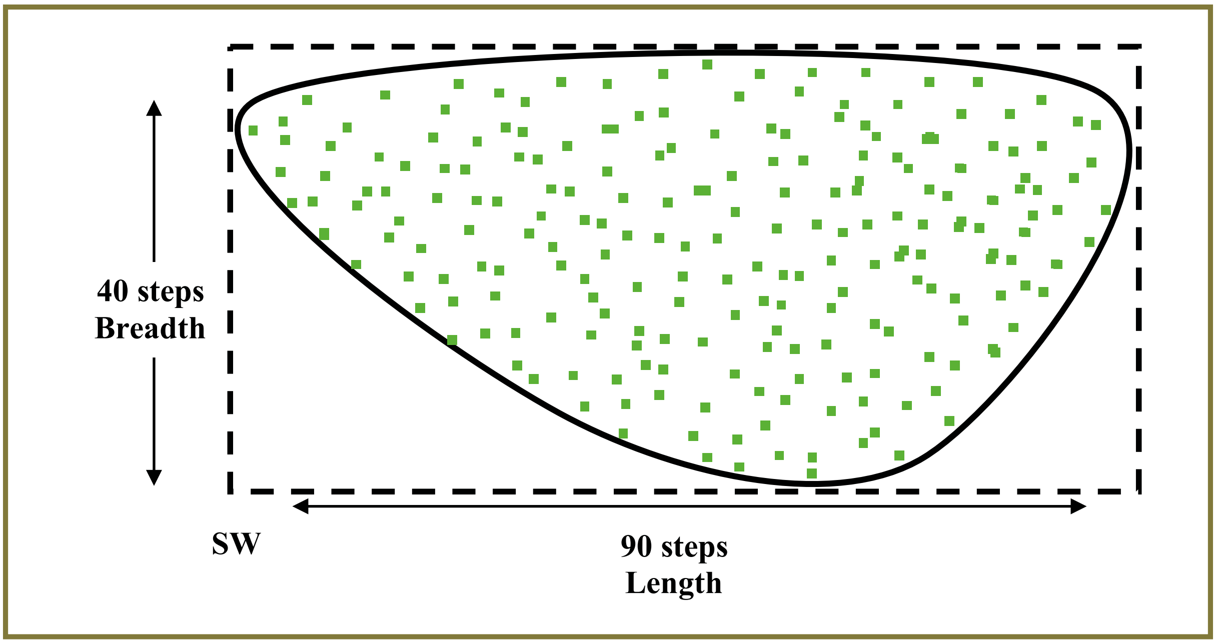
As a rule of thumb, locating an experimental plot within a selected field requires first, identifying the southwest (SW) corner of the field using a compass or GPS. Identifying the SW corner of regularly shaped fields are not as difficult as irregular shaped fields. However, to avoid any inconvenience, an irregularly shaped field is usually enclosed in a regular shape by outer straight line measurements (Sud et al., 2015). The longer and shorter sides of the field are usually referred to as length and breadth, respectively.

FIGURE A6.2. REGULARLY SHAPED FIELD.



Source: Sud *et al.*, 2015.

FIGURE A6.3. IRREGULARLY SHAPED FIELD.



Source: Sud *et al.*, 2015.

Siting one rectangular experimental plot. Suppose that the survey objective is to randomly site one rectangular plot of size 10 m x 5 m in each selected field. It should be noted that the random location of the experimental plot and the harvesting of the crop are carried out on the same day. In this example, the regularly shaped field (Figure A6.3) is considered as the selected field. After identifying the SW corner of the field using the compass or GPS, the next task is to deduct 14 steps and 7 steps from the length and breadth of the selected field, respectively, to ensure that the rectangular plot of size 10 m x 5 m properly fits in the selected field. Deducting 14 steps from the length of 100 steps gives 86 steps, while deducting 7 steps from the breadth of 50 steps gives 43 steps.

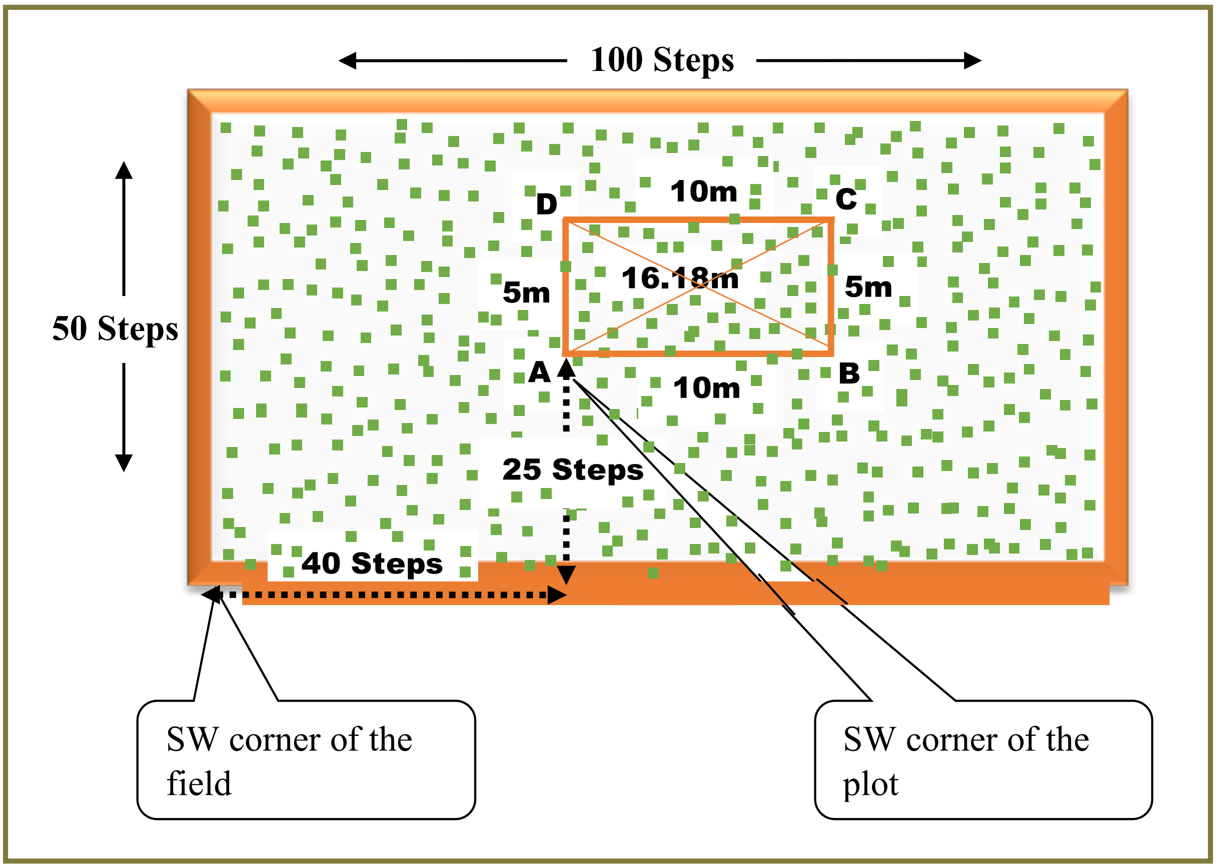
Using a typical random number table (note: special random number tables are usually distributed to enumerators), columns 1 and 2 are assigned to the length and breadth, respectively. Thus, only the first two digits in each column will be considered because after the subtractions, the length and breadth comprise only of two-digit numbers.

Suppose that the random table produced 40 for length and 25 for breadth. It is important to note that if the random number generated will make the experimental plot or even part of the plot exceed the boundary of the selected field, then a new random number should be selected or generated. The intersection of the pair or simply the pair (40, 25) indicates the SW corner of the rectangular plot of size 10 m x 5 m in the selected field. Starting from the SW corner of the selected field, measure 40 steps along the length of the field and then at this point, measure 25 steps by entering into the field, that is, vertical to the length but parallel to the breadth of the field. Let the point of the pair (40, 25) be indicated as point "A".

Place a long peg at point "A". Now, from point "A", measure 10 m parallel to the length of the field and indicate it as point "B". Fix another long peg at point "B". This is the second corner. For crops that are short such as rice, wheat and groundnut, applying a right-angle triangle method in marking the third ("C") and fourth ("D") points will be appropriate. However, for long crops such as cassava and maize, the use of a compass or GPS in marking the points "C" and "D" is recommended to avoid any inconvenience.

In applying the right-angle triangle method, there is the need to compute the distance of the diagonal, "AC" = "BD", which is $\sqrt{(length)^2 + (breadth)^2} = \sqrt{[(10)^2 + (5)^2]} = 11.18$. This means that the distance from point "A" to "C" to "B" or "ACB" = "AC" + "BC" = 11.18 + 5 = 16.18. Now, to obtain point "C", three persons are required. Let one person stand at point "B" and another person at point "A" holding the measuring tape at the zero-meter (0 m mark and 16.18 m mark, respectively. The third person should stretch the tape and fix the point "C" in such a way that the distance "AC" is 11.8 m and "BC" is 5 m. Likewise, point "D" is fixed by allowing one person to stand at point "A" and another person at point "B" holding the measuring tape at the 0 m mark and 16.18 m mark, respectively. The third person should stretch the tape and fix the point "D" such that the distance "AD" is 5 m and "BD" is 11.18 m. Make sure that long pegs are erected at points "C" and "D" as well. The points "ABCD" constitute the four corners of the rectangular experimental plot of size 10 m x 5 m. Finally, use the measuring tape to check whether indeed all the distances, such as "AB" = "CD" = 10 m, "AD" = "BC" = 5 m and "AC" = "BD" = 11.18 m were correctly measured.

FIGURE A6.4. RECTANGULAR EXPERIMENT PLOT OF SIZE 10 M X 5 M DEMARCATION.



Source: Sud *et al.*, 2015.

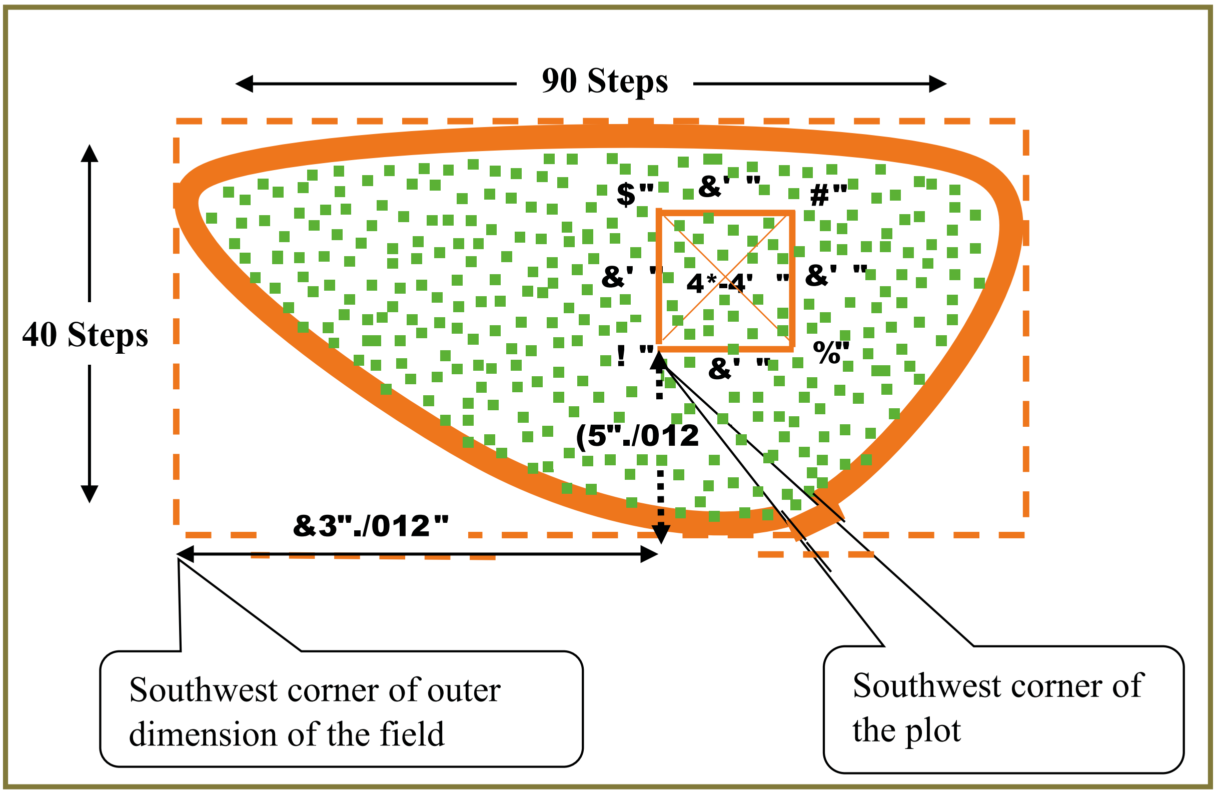
Siting one square experimental plot: consider that in this example, the survey objective is to randomly locate one square plot of size 5 m x 5 m in each selected field. The steps for siting a square experimental plot and a rectangular experimental plot are alike, except that because of the difference in the lengths, appropriate adjustments are made to ensure that the location of the square plot is feasible. For example, let the irregularly shaped field (Figure A6.4) be the selected field.

Use a compass or GPS to identify the SW corner of the field and deduct seven steps from both the length and the breadth. Thus, after the deduction, the new length is 83 steps and the new breadth is 33 steps. Suppose that the random number pair generated for both length and breadth are 52 and 19, respectively. Ensure that the random numbers generated or selected would not make the square plot exceed the boundary of the field, otherwise new random numbers should be generated until siting of the square plot would be within the boundaries of the field.

The random number pair (52, 19) shows the SW corner of the experimental plot (for example, a square plot of size 5 m x 5 m) in the selected field. Using the same steps that were followed in locating the rectangular shape discussed above, fix long pegs at points "A" and "B" after locating them. The distance of the diagonal, "AC" = "BD", is computed as $\sqrt{(length)^2 + (breadth)^2} = \sqrt{(5^2 + 5^2)} = 7.07$.

The distance from point "A" to "C" to "B" or "ACB" is calculated as AC + BC = 7.07 + 5 = 12.07. If the crops are short, right-angle triangle method should be applied to obtain the points "C" and "D". However, if the crops are long, then the use of a compass or GPS is appropriate. Long pegs should be fixed at points "C" and "D" after locating them. Finally, use the measuring tape to check the accuracy of all the dimensions, such as "AB" = "CD" = "AD" = "BC" = 5 m and "AC" = "BD" = 7.07 m.

FIGURE A6.5. SQUARE EXPERIMENT PLOT OF SIZE 5 M X 5 M DEMARCATION.



Source: Sud *et al.*, 2015.

a. **Harvesting the experimental plot and other operations:**
The farm holder is expected to wait until the experimental plot is harvested before harvesting the rest of the field or farm. To ensure that only crops planted within the located plot are harvested, a string should be tied firmly around each of the long pegs positioned at the four corners of the plot.

For some crops, such as rice and wheat it will be appropriate to lower the string down to the ground level to clearly demarcate the boundary of the plot in such a way that a clear distinction can be made between a crop that falls within the plot and a crop that is outside the plot.

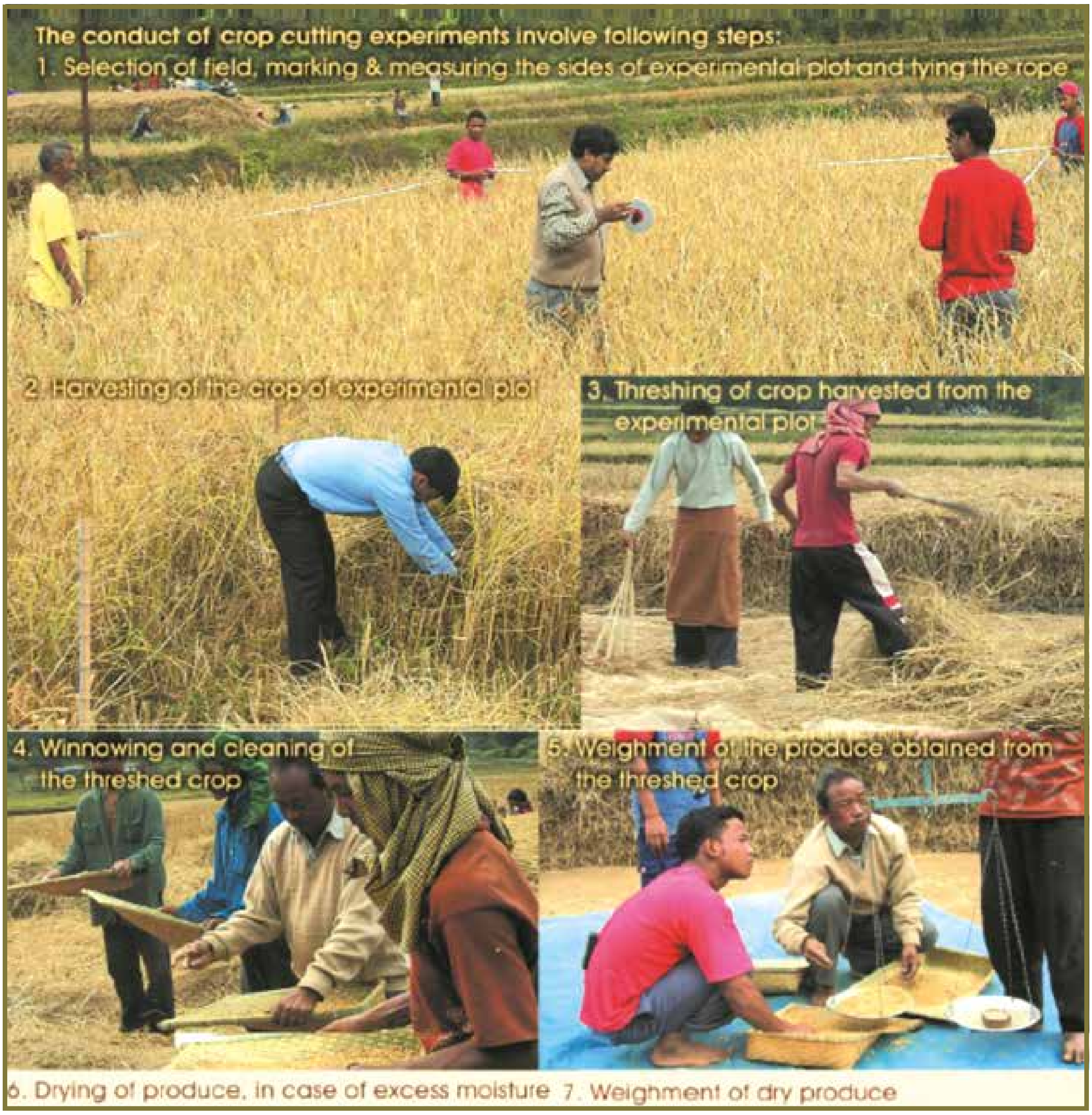
The rule is that if a crop lies on the boundary of the plot, then it is considered as part of the crops within the experimental plot to be harvested if and only if more than half of its roots fall inside the plot, otherwise it is rejected.

The crop cutting method is expected to measure biological yield (that is, yield without considering any loss); thus, activities such as shattering of grain in the plot during harvest as well as leaving ear heads in the plot, which usually lead to harvest losses, should be avoided. The harvested crop should be carried from the experimental plot to the threshing ground or place. The harvested crop from each experimental plot must be kept separate. Strict adherence to appropriate and recommended threshing and winnowing procedures regarding different types of crops should be followed, to guarantee dust-free grain or produce devoid of any foreign material contamination or clean produce. The clean produce should be weighed accurately, even though it may contain a certain amount of moisture component. All weights should be recorded as expected.

b. **Final weighing after further drying**
The produce of some crops such as rice and maize require further drying before the final drying weight can be obtained for estimation of the final crop yield. Typically, a random sample of the harvested clean produce, usually 1 kg, is collected in a bag or sack and dried using the recommended drying method or in sunlight for a certain number of days. After the required number of days, the dried produce is weighed to obtain the final dry weight for crop yield estimation.

Figure A6.6 shows the various steps involved in a crop cutting operation.

FIGURE A6.6. STEPS INVOLVED IN THE CROP CUTTING METHOD.



Source: Sud et al., 2015

Annex 7

The crop area, production and yield estimation system in the United States of America

The method for estimating crop production and yield followed in the United States of America involves a full range of data collection and estimation methods, including sample surveys, agricultural censuses, administrative data, remote sensing and crop modelling. The United States National Agricultural Statistics Service (NASS) of the USDA publishes 173 crop reports annually. NASS has an annual estimating program that accounts for 95 percent of production and that is reviewed following each CA. NASS estimates planting intentions, planted area, and throughout the growing season, it forecasts yield and production of crops. At the end of the cropping season, NASS estimates final acreage, yield and production. Throughout the growing season, NASS reports crop progress and condition ratings every week.

NASS conducts quarterly agricultural surveys, an annual area survey, and monthly agricultural and objective yield surveys (Miller and Hoffman, 2016). In March, NASS surveys approximately 86 000 farmers on their planting intentions. The June Agricultural Survey (JAS) covers 73 000 farmers and approximately 11 000 segments (of the area frame) to collect data on crop area planted and intended for harvest. In September, around 66 500 small grains farmers are surveyed while in December, about 85 000 farmers that plant row crops are surveyed for final area harvested, yield and production. All these quarterly surveys also collect data on the grains stored on the farm.

JAS has a multiple-frame sampling design. There are basically two stages of land classification in the area frame. First, it is divided according to its utilization or cultivation, that is, from intensively cultivated areas to marginally cultivated grazing areas to urban areas. In the second stage, the utilization classification in the first stage is subdivided into segments, usually ranging from 1 square mile in intensively cultivated areas to 0.1 square miles in urban areas, through the use of satellite imagery, map products and other tools (Vogel and Bange, 1999). The list frame consists of the list of farmers in the United States of America. It is deemed incomplete because farm operations could go into and out of business any time. The JAS is also used to update the list frame (Davies, 2009).

In the JAS area frame survey, data collectors visit each sample segment that has been accurately identified on aerial photography and interview all holders that operate land inside the boundaries of the selected segment.

The JAS is also administered to another random sample of farmers that is obtained from the list frame. Data collection is done through telephone interviews.

To avoid an overlap of the area and list frames, NASS has developed an efficient procedure of identifying duplicate sample farmers and making sure that they appear only in the list frame. In the rare cases that a sample farmer is surveyed twice, his or her response will only be counted once (Vogel and Bange, 1999).

Follow-up monthly surveys are conducted within the planting season using a subset of sample farmers from the list frame used in the March agricultural survey or the JAS to estimate the area that will be harvested, expected crop yield and actual crop yield obtained at harvest. These agricultural yield surveys are conducted monthly from May to November. The crops covered vary monthly and the sample sizes range from 6 000 to 28 000. Farmers report expected yields for their crops as of the condition on the first day of the month of the survey.

Objective yield surveys are also conducted from May to December only for corn, soybean, winter wheat and cotton. Samples sizes range from 1 217 to 1 835 and were selected from the area frame used in the March or June area frame survey. Objective yield surveys are only conducted in the largest producing states accounting for 75 percent of agricultural land in the United States of America (Miller and Hoffman, 2016). Trained data collectors visit the sample fields to observe, measure and collect samples that they send to the laboratory for analysis. During the early stages of growth of the crops, plant characteristics such as colour, plant count or density and plant vigour are used as prediction variables for estimating crop maturity and yield forecasts. At the ripened stage of the crops, appropriate objective measurement method is applied to obtain harvest estimates such as counts, measurements and other observations from each sample plot. These harvest estimates are used as input variables in statistical models based on historical data to obtain final estimates such as number of fruit and weight per fruit (Vogel and Bange, 1999; Aune *et al*., 2006).

NASS uses various administrative data to improve the estimates derived from the surveys mentioned above. For example, the estimates of area planted derived from the March agricultural survey are revised based on the reports from an USDA production support programme, in which most of the field crop farmers are enrolled. Data from millers, cotton gins and traders are also considered. A balance sheet in which production, consumption, export and import of a crop is constructed to ensure consistency of these data (Nusser and House, 2009; Aune *et al*., 2006).

Final harvested yield and area planted are obtained by surveying the farmers who were involved in the objective measurement surveys through telephone interviews, at the end of the harvest period. The harvest loss estimate is obtained by revisiting the sample fields, and grain left on the ground in each selected plot is gathered and weighed as a measure of loss (Vogel and Bange, 1999).

The system for estimating crop area, production and yield is comprehensive, timely and reliable. It is based on well-established methods that are continually improved through methodological research. Developing countries could study the system and adapt some of its components. However, taking the system as a whole requires resources and expertise generally beyond the limits of what developing countries could afford.

Annex 8

Mixed cropping and area apportioning under mixture: when to do it?

A significant portion of farmers in developing countries, especially smallholders, are involved in the practice of polyculture systems, which consist in the presence of plots under mixed cropping.

Table A8.1 below presented the results of three nationwide surveys conducted under the West Africa Agricultural Productivity Program (WAAPP[1]). Using data collected through these surveys, the proportion of plots under mixed stand was computed for Guinea, Mali and Côte d'Ivoire.

TABLE A8.1. PROPORTION OF PLOTS UNDER MIXED STAND FROM A SAMPLE OF PLOTS SELECTED AT NATIONAL LEVEL UNDER THE WAAPP BASELINE SURVEY.

		Guinea (Plots under rice in mixture with other crops) 2013/2014	Mali (Any crop) 2015/2016	Côte d'Ivoire (Any crop) 2015/2016
Number of plots visited		1 649	1 970	6 074
Proportion of plots under mixed stand (%)		57.3	48.3	33.2
Proportion of plots under mixed stand disaggregated by number of crops in the mixture (%)	2	19.9	25.4	22.5
	3	11.6	13.3	7
	> 3	25.8	9.7	3.7

Sources of data: Agricultural Surveys ENSEA/WAAPP, 2013/2014, 2015/2016.

1 http://projects.worldbank.org/P094084/west-africa-agricultural-productivity-program-waapp?lang=en

Various types of mixed stand patterns are possible, and the presence of different crops with different characteristics and growing cycles poses various challenges in the computation of statistics on crop area, production and yield.

The various types of mixed crop patterns are schematized below:
- **Intercropping**: examples are plantain intercropping in cocoa plantations and, yam in a cashew plantation
- **Crops along the borders of the plot of the main crop**: horticulture practice along the border of a cotton farm, cocoa plantation, palm tree farm, cashew plantation, etc.
- **Mixed stand with a homogeneous mixture of the plants or bunches on the parcel**
- **Seedlings spread out**: seedlings of maize or plantain spread in a yam field, watermelon spread out in maize field, etc.
- **Crops sown in spots (in the mixture)**: few palm trees in a cocoa plantation, plantain in spots in cocoa plantation, yam in spots in cashew plantation, etc.

This annex is an attempt to answer the following questions:
- Is it always advisable to allocate areas under crop in case of mixture?
- What are the criteria to be considered in deciding whether a given scenario is eligible for area allocation?
- In which cases can the physical area sown be equivalent to the area allocated to each crop in the mixture?

In order to answer these questions, some examples of mixed stands are provided:
Mixed Stand 1: Plantain (as shade tree) in cocoa plantation (for the first two years of the cocoa plantation) (intercropping).
Mixed Stand 2: Cashew plantation in mixture with maize, legumes (peanut, cowpea, soybean, pigeon pea), cassava, yam, etc.
Mixed Stand 3: Rain-fed rice and maize, etc.
Mixed Stand 4: Maize in mixture with groundnut, watermelon, cucurbits, pumpkins, etc.

Some of the above mixtures can be sown in compliance with the recommended density (in pure or mixed stand) and they grow together even if the sown time could differ in terms of compliance with the crop calendar. The main criteria to be considered in deciding whether to do apportioning or not are the following:
- Are the two crops in competition for growth?
- Are the densities in compliance with the one recommended in pure stand (or sole crop)?

No competition for growth

Coulibaly et al. (2012d) show that the association of legumes with maize results in a decrease in maize grain yield compared to pure stand, because of competition problems between cereals and legumes and the density applied. Indeed, density greater than that recommended for each crop in mixed stand may create competition between crops. Below are a few examples of this scenario:

Case 1. Rice or groundnut could be in concurrence (competition) with maize. On the other hand, maize could be in competition with both of them if the maize density exceeds that recommended in mixed stand. Thus, for rice or groundnut, the density can be applied, while for maize, it is important to avoid its dominance over the recommended density in mixture. The opposite could be observed too: recommended density for maize, while for rice or other legumes, the density should be lower than that in sole crop.

Thus:

- if the density is respected for one of the crops (rice, groundnut or maize), there is no need to apply any apportioning approach;
- the area under maize (or rice or groundnut, in the second scenario) needs to be evaluated based on the density applied.

The same observation could be made for maize and watermelon in a mixture. Indeed, maize suffers less yield loss when intercropped with watermelon, so that its density in sole crop could be followed (Munisse et al., 2012), while for watermelon, portioning should be applied according to the density or the mixture type.

Case 2. In the case of the cocoa plantation, plantain plays an important role of shade tree to protect the cocoa plants. In that case, the plantain's shade is advantageous to the growth of the fragile young cacao plants. Therefore, both crops grow together, under the recommended density in pure stand without any competition up to the time the cacao starts dominating over the plantain tree (after two years).

Recommended density and mixed stand

Insofar as the density in the mixture for each crops complies with that recommended in pure stand, the physical area of the plot is equivalent to the area sown (or harvested) for each crop. The examples below are illustrative.

Mixed stand 1: plantain and cocoa tree

The density recommended in pure stand by the CNRA in Côte d'Ivoire:

- for cacao, 1 333 cocoa trees/ha (3 m between the lines and 2.5 m between the seedlings on the lines)
- for cacao, is 1 666 plantain trees/ha (3 m between the lines and 2 m between the seedling on the line). Sometimes, the same density is also recommended: one cocoa tree o one plantain[2].

The same density is also recommended in the case of mixed stand for both crops together.

In that case, the densities are respected. The young cocoa plant would not compromise the growth and the production of the plantain for the first two years, and vice versa. After two years, the cocoa takes the lead. In that case, for the first two years, the area under each crop is equivalent the physical area of the plot.

Mixed Stand 3: rice and maize

The different confirmation for rain-fed rice recommended in pure stand by CNRA in Côte d'Ivoire:

- For rain-fed rice, 160 000 bunches/ha is recommended in pure stand;
- For example, in mixture rainfed rice and maize, this configuration is also possible: a density per hectare of 5 000 maize plants and 160 000 bunches of Nerica 1 (rain-fed rice) and the equivalent area ratio of 1.21 for the associated crops. Indeed, that is the maize density that enables rice to achieve the best yield: one line of maize for seven lines of rice spaced 25 cm apart.
- However, in pure stand, the density of maize varies from 53 000 plants/ha to 62 500 pieds/ha.

Example

To obtain a real estimate of cocoa production, the Ministry of Agriculture decides to carry out a data collection operation for one specific year at national level. The information required relates to the land area under cocoa threes and the yield by agro-ecological area. The enumerators are asked to measure the area of cocoa plots and lay down yield and density grids. The presence of cocoa plantation under mixed or intercropping is widespread.

The enumerators in the field found the following scenarios:

- Event 1: a new plot planted with cocoa (1 333 seedlings/ha, which is the optimum density for this type of plant). On the same plantation, plantain trees have been used to provide shade to the cocoa.
- Event 2: a cocoa plantation in production has a density of 1 333 seedlings/ha. Yam bulbs have been planted between the rows (intercropping).
- Event 3: a yam plot in which cocoa plants have been dispersed.
- Event 4: a palm oil plot with a few cocoa plants spread around. The cocoa is estimated to take up less than 10 percent of the surface area.

Questions

Q1. Which cases should be taken into account to assess area under cocoa?

Q2. Which cases should be taken into account to assess production?

Q3. For each case above, explain how the area under crop should be distributed.

2 "The density of a banana tree for a cocoa tree (1333 plants / ha) planted the same year in the interlinings, gives the best results and is the subject of recommendation" (Lachenaud, 1987). Available at https://nationalzoo.si.edu/scbi/migratorybirds/research/cacao/koffi2.cfm

Solution

C1.The various events (1 to 4) can be considered in assessing production area for cocoa. However, in case 3 (and possibly case 4), the area allocated to cocoa may be marginal. It is therefore a trade-off between the additional information gathered (in cases 3 and 4) and related costs. In those cases:

a. if the cocoa trees are used just for shelter (from sun or rain) or for workers to rest under (rather than for production purposes), the area in that case must be ignored
b. if the farmer states that the production – albeit marginal – is to sell, then it is advisable to count the number of trees spread over the plot and derive the surface area by dividing the number of seedling trees by the density

C2.The case that are not to be taken into account to assess production is 1. Indeed, even if case 1 was taken into account to assess area under crop, it cannot be considered for production purposes has the plantation has not yet reached production stage – it is in the growth phase. Furthermore, in production assessments, it is useful to consider the age of the plantation to categorize the farm under one of the following headings: growth phase, production phase, rated output and decline. This classification is necessary to properly ascertain yield. A 30-year-old plantation will have started to decline.

C3.Portioning of area under cocoa and other crops:

- In cases 1 and 2, there is no portioning to carry out. Indeed, the recommended density for cocoa was respected. Furthermore, the presence of yams is not in competition with cocoa that has reached its production phase. On the other hand, assessing the area under yams should take the density into account.

Conclusion

In mixed cropping scenarios, special attention should be paid in the production of agricultural statistics regarding the estimation of areas planted, yield and production.

In some cases, it is advisable to conduct apportioning, and to refrain in other cases. The criteria to consider before any step to area apportioning are:

- type of mixture (intercropping, spots seedling, crop on border, etc.);
- density of standing crops (in compliance with density in pure stand); and
- the fact that crops should not suffer yield (production) loss because of the competition in the mixture.

As long as the density applied in mixture is the one recommended in pure stand, and as long as there is no competition and the crops are homogeneously scattered (or there is intercropping), the total physical area should be allocated to each crop. In other scenarios, such as crops in border or in spot, the area need not be considered if the share occupied is insignificant.

Annex 9

Tables from chapter 7

TABLE A9.1. 2014 CASSAVA PRODUCTION IN ZANZIBAR (KG).

Month	D1		D2		R1	R2	Crop Cut
June	209.74		223.32				
July	112.4		138.61				
August	118.93	715.45	120.16	854.1	734.38		
September	97.69		148.41				
October	82.18		110.49				
November	94.51		113.11				
December	107.25		111.55			1103.2	2074.8
January	67.63		125.62				
February	98.46	522.95	101.95	719.73	573.0		
March	80.01		130.29				
April	83.28		115.27				
May	86.32		135.05				
TOTAL	**1238**		**1574**		**1307**	**1103**	**2075**

Source: Author elaboration, 2018.

TABLE A9.2. SELECTED COEFFICIENTS FROM PRODUCTION AND YIELD REGRESSION IN MALAWI.

	Total Production (Kg)		Yield (Kg/Ha, GPS)		Yield (Kg/Ha, GPS)	
	Mean	Coefficient	Mean	Coefficient	Mean	Coefficient#
Diary - Visit †	1,072	N/A	5,208	N/A	5,208	-3582 (507)
Diary - Phone †	1,391	295*** (80)	6,618	1431*** (430)	7,717	-2211*** (591)
6-Month Recall †	1,102	37 (68)	5,798	561 (400)	5,798	-2990*** (434)
12-Month Recall †	844	-221*** (61)	4,671	-617*** (337)	4,671	-4187*** (444)
Comparison Category	Diary - Visit †		Diary - Visit †		Crop Cutting	
Comparison Category Mean	**1072**		**5208**		**8958**	
Controls Included?	YES		YES		YES	
Observations	1,218		1,218		2345	
R2	0.45		0.36		0.44	
Tests of Equality of Coefficients						
D1 = D2	--		--		0.00	
D1 = R1	--		--		0.35	
D1 = R2	--		--		0.03	
D2 = R1	0.00		0.04		0.06	
D2 = R2	0.00		0.00		0.00	
R1 = R2	0.00		0.01		0.01	

Notes: † denotes a dummy variable. Constant estimated but not reported. ***/**/* denote statistical significance at the 1/5/10 percent level, respectively. ‡ denotes standard errors clustered at the enumeration area-level. # denotes standard errors clustered at the household-level.

Source: Author elaboration, 2018.

TABLE A9.3. IMPLEMENTATION COST ESTIMATES (USD) BY TREATMENT ARM.

	Diary - Visit	Diary - Phone	6-Month Recall	12-Month Recall
Fixed Fieldwork Costs				
Survey & Questionnaire Design	667	667	667	667
Equipment	7,496	7,496	7,496	7,496
Household Listing	7,069	7,069	7,069	7,069
Training	11,527	11,527	11,527	11,527
Local Presence & Incentives	11,303	11,303	11,303	11,303
Variable Fieldwork Costs				
Per Diems & Salaries for Field Staff	60,429	24,429	10,072	5,036
Per Diems & Salaries for Management	11,914	4,816	1,986	993
Fuel & Maintenance for Field Teams	5,592			
Fuel & Maintenace for Management	19,965	8,071	3,328	1,664
Fuel to Travel to field	1,888	4,879	2,252	1,126
Airtime for Field Staff	2,228	901	371	186
Airtime for Management	1,308	529	218	109
Incentives				
Weighing Scale	3,051	3,051		
Sacks	1,145	1,145		
Mobile Phone		4,950		
Solar Charger		4,050		
Airtime		2,160		
Cash	2,250	1,125	2,250	2,250
Call Center				
Facility/ Equipment		36		
Staff		2,058		
Airtime		3,771		
Total Cost	147,832	104,033	58,538	49,425
Cost Per Household	469	330	186	157

Source: Author elaboration, 2018.

Annex 10

Journal for vegetable crop surveys

Journal, page 1 beds areas

Version 1, May 2018

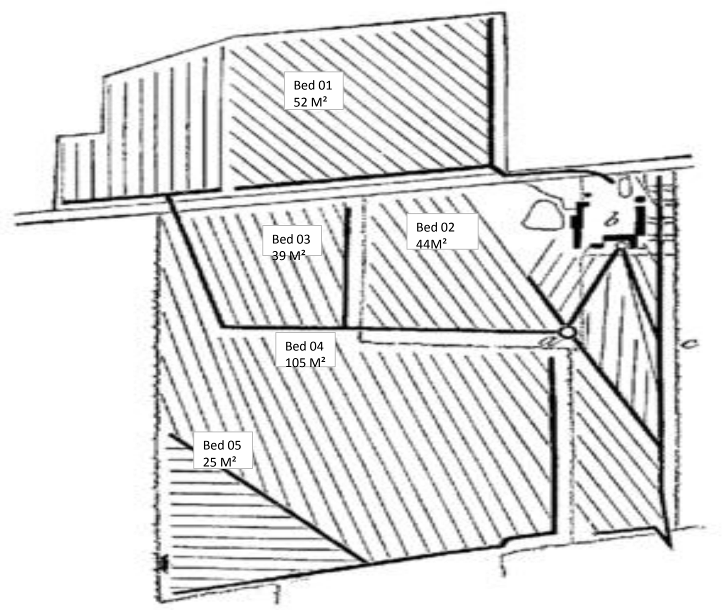
This model of journal is proposed to help respondents to register their activity during the reference period in order to fill in the questionnaire or module proposed for vegetables and ornamental plants. It must be left to the respondent at the beginning of the reference period and compiled by the surveyor at the end of the period, in order to fill in the questionnaire that must be validated by the respondent. Between those two visits, the surveyor should contact the respondent to answer his (her) questions and remind him (her) to regularly fill in the journal.

Reference period: DD/MM/YYYY to DD/MM/YYYY

Surveyor first name: ☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐ **Surname:** ☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐ **Surveyor number:** ☐☐☐☐☐☐

Surveyor telephone number: ☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐

MAP OF THE BEDS USED FOR HORTICULTURAL CROPS



LIST OF BEDS USED:

	Unit of measure	Conversion factor to a standard unit
What is the area unit used throughout the journal?	☐☐☐	☐☐☐☐☐☐

	Area	Under greenhouse	Hydroponic technique	Low shelters (possible)	Always open field	Irrigated	Drained	Under organic farming
BED 01	☐☐☐☐☐.☐☐	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes
BED 02	☐☐☐☐☐.☐☐	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes
BED 03	☐☐☐☐☐.☐☐	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes
BED 04	☐☐☐☐☐.☐☐	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes
BED 05	☐☐☐☐☐.☐☐	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes
BED 06	☐☐☐☐☐.☐☐	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes
BED 07	☐☐☐☐☐.☐☐	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes
BED 08	☐☐☐☐☐.☐☐	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes
BED 09	☐☐☐☐☐.☐☐	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes
BED 10 …	☐☐☐☐☐.☐☐	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes	◯ 0 No ◯ 1 Yes

Journal, fertilizing

Version 1, May 2018

All activities should be recorded daily in order not to forget operations during the reference period. At the end of the period, it will be used to fill in the questionnaire that must be validated by the respondent. Between those two visits, the surveyor should contact the respondent to answer his (her) questions and remind him (her) to regularly fill in the journal.

Reference period: DD/MM/YYYY to DD/MM/YYYY

Date (DD/MM/YYYY)	Bed number	Part of the bed fertilized	Fertilizers applied						Comments
			Mineral	Organo-mineral	Compost	Mulch	Bio	animal effluents	
		%	○	○	○	○	○	○	
		%	○	○	○	○	○	○	
		%	○	○	○	○	○	○	
		%	○	○	○	○	○	○	
		%	○	○	○	○	○	○	
		%	○	○	○	○	○	○	
		%	○	○	○	○	○	○	

...

Journal, sowing, planting

Version 1, May 2018

All activities should be recorded daily in order not to forget operations during the reference period. At the end of the period, it will be used to fill in the questionnaire that must be validated by the respondent. Between those two visits, the surveyor should contact the respondent to answer his (her) questions and remind him (her) to regularlyfill in the journal.

Reference period: DD/MM/YYYY to DD/MM/YYYY

*** Report all vegetables and ornamental plants sown or planted.**
*** Report associated crops grown on the same bed.**
*** See vegetables and ornamental plants list.**

Date (DD/MM/YYYY)	Vegetable or ornamental plant sown or planted		Bed number	Part of the bed sown or planted	Comments
	Name	Code			
				%	
				%	
				%	
				%	
				%	
				%	
				%	

...

Journal, plant protection products (PPPs)

Version 1, May 2018

All activities should be recorded daily in order not to forget operations during the reference period. At the end of the period, it will be used to fill in the questionnaire that must be validated by the respondent. Between those two visits, the surveyor should contact the respondent to answer his (her) questions and remind him (her) to regularly fill in the journal.

Reference period: DD/MM/YYYY to DD/MM/YYYY

Date (DD/MM/YYYY)	Bed number	Part of the bed treated	Objective of the PPP applied					Comments
			Insecti-cide	Herbi-cide	Fungi-cide	Rodenti-cide	Other	
		%	○	○	○	○	○	
		%	○	○	○	○	○	
		%	○	○	○	○	○	
		%	○	○	○	○	○	
		%	○	○	○	○	○	
		%	○	○	○	○	○	
		%	○	○	○	○	○	

...

Journal, harvesting

Version 1, May 2018

All activities should be recorded daily in order not to forget operations during the reference period. At the end of the period, it will be used to fill in the questionnaire that must be validated by the respondent. Between those two visits, the surveyor should contact the respondent to answer his (her) questions and remind him (her) to regularly fill in the journal.

Reference period: DD/MM/YYYY to DD/MM/YYYY

*** Report all vegetables and ornamental plants harvested.**
*** Report associated crops grown on the same bed.**
*** See vegetables and ornamental plants list.**

Date (DD/MM/YYYY)	Vegetable or ornamental plant harvested – Name	Vegetable or ornamental plant harvested – Code	Bed number	Part of the bed sown or planted	Quantity harvested	Unit of measure (see codes)
				%		
				%		
				%		
				%		
				%		
				%		
				%		

…

Version 1, May 2018

All activities should be recorded daily in order not to forget operations during the reference period. At the end of the period, it will be used to fill in the questionnaire that must be validated by the respondent. Between those two visits, the surveyor should contact the respondent to answer his (her) questions and remind him (her) to regularly fill in the journal.

Reference period: DD/MM/YYYY to DD/MM/YYYY

*** Report all vegetables and ornamental plants that leave the holding and describe the different uses.**
*** See vegetables and ornamental plants list.**

Date (DD/MM/YYYY)	Vegetable or ornamental plant leaving the holding: Name	Code	Quantity leaving the holding	Unit of measure (see codes)	Self consumption: Quantity	or	Percentage
						or	%

Used as pay or wages for labour: Quantity	or	Percentage	Given to service or input providers: Quantity	or	Percentage	Used for on-farm processing: Quantity	or	Percentage	Losses after harvest: Quantity	or	Percentage
	or	%		or	%		or	%		or	%

Direct sale to consumers: Quantity	or	Percentage	Total amount	or	Unit price	Sold to industry, wholesaler or reseller: Quantity	or	Percentage	Total amount	or	Unit price
	or	%		or			or	%		or	

Date (DD/MM/YYYY)	Vegetable or ornamental plant leaving the holding: Name	Code	Quantity leaving the holding	Unit of measure (see codes)	Self consumption: Quantity	or	Percentage
						or	%

Used as pay or wages for labour: Quantity	or	Percentage	Given to service or input providers: Quantity	or	Percentage	Used for on-farm processing: Quantity	or	Percentage	Losses after harvest: Quantity	or	Percentage
	or	%		or	%		or	%		or	%

Direct sale to consumers: Quantity	or	Percentage	Total amount	or	Unit price	Sold to industry, wholesaler or reseller: Quantity	or	Percentage	Total amount	or	Unit price
	or	%		or			or	%		or	

Date (DD/MM/YYYY)	Vegetable or ornamental plant leaving the holding: Name	Code	Quantity leaving the holding	Unit of measure (see codes)	Self-consumption: Quantity	or	Percentage
						or	%

Used as pay or wages for labour: Quantity	or	Percentage	Given to service or input providers: Quantity	or	Percentage	Used for on-farm processing: Quantity	or	Percentage	Losses after harvest: Quantity	or	Percentage
	or	%		or	%		or	%		or	%

Direct sale to consumers: Quantity	or	Percentage	Total amount	or	Unit price	Sold to industry, wholesaler or reseller: Quantity	or	Percentage	Total amount	or	Unit price
	or	%		or			or	%		or	

...

Annex 11

Model for questionnaires for vegetable crop surveys

Questionnaire for vegetable crops

Version 1, May 2018

This model of questionnaire is proposed as an independent annual survey. It is based on AGRIS questionnaires, with the objective of describing the target activity as well as its context. The questionnaire should be adapted or simplified following the individual country's reality and needs. If the horticultural crops production survey is administered as a module depending on a main survey, sections 1 and 2 proposed here can be skipped as they are part of the main survey. If necessary, the main survey should be adapted if some data are missing.
Surveying horticultural plants production requires at least two visits. The first visit, at the beginning of the reference period, will be devoted to the description of the holding and the capability of the respondent to answer questions about the holding activity at the end of the period. If this activity is not recorded in sufficient detail, a ***journal will be left and explained to the holder*** *for him or her to fill in. In this case, at the end of the period, the main questionnaire will be completed by the surveyor and approved by the holder.*

SECTION 1: THE HOLDING

PART 1.1: SURVEY PREPARATION

ID code of the holding
Name and surname of the holder or legal name of the holding
Legal status of the holding
Address: region, district, town, enumeration area, legal address, phone number, type of address (plot, house, agricultural building)
GPS coordinates: longitude, latitude

Surveyor first name: □□□□□□□□□□□□□□□ **Surname:** □□□□□□□□□□□□□□□ **Surveyor number:** □□□□□□

Start time of the survey: □□ hour □□ minutes

Questions Q01 and Q02 below will be answered by the surveyor.

Q01. Did I find a farm at the same address or same name?
◯ 0 No → **Go to Section 4. Explain in the comments box how you tried to find the holding and stop the questionnaire.**
◯ 1 Yes

Q02. Did I find somebody from the holding who accepted to answer?
◯ 0 No → **Go to Section 4. Explain in the comments box how you tried to obtain acceptance and why the person does not want to answer, and stop the questionnaire.**
◯ 1 Yes → **Introduce the survey using the text below:**

TEXT TO READ:
I represent the [SURVEY INSTITUTE]. We are conducting a survey about horticultural production in all regions of [COUNTRY].
You are aware that people talk a lot about the current situation of agriculture in [COUNTRY]. To have an exact idea of what is really going on, there is no other way than to survey the conditions of thousands of holdings in various categories. Only after summarizing all of these responses can we have a real picture of the horticulture of [COUNTRY].
For that purpose, [number of] holdings have been chosen at random as in a lottery. One of these holdings proved to be yours. The authenticity of the results of the whole survey will depend on your sincerity and exactness when answering the questions of this questionnaire on the activities and production of your holding.
We assure you that your personal responses will not be disclosed and after all these questionnaires are processed by a computer, they will be used only in a summary way.
If you have any questions in connection with this survey, you are welcome to telephone the number indicated on the visit card of our organization that I leave for you here.
I express my gratitude for your participation in this survey in advance.

Q03. Record the following information about the respondent.
Q03a First name . . . □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
Q03b Surname . . . □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

Q03c Sex
◯ 1 Male
◯ 2 Female

Q03d What is your function on the agricultural holding?

(Fill in one circle only)
◯ 1 Holder (legal and/or economically responsible for the holding) → **Explain the detailed survey objectives using the text below.**
◯ 2 Co-holder (legal and/or economically co-responsible for the holding) → **Explain the detailed survey objectives using the text below.**
◯ 3 Manager (responsible for the day-to-day decisions on the farming operations) → **Explain the detailed survey objectives using the text below.**
◯ 4 Employee or household member working on the holding → **Explain the detailed survey objectives using the text below, then ask the following question:**

Are you able to answer questions for the agricultural holding?
◯ 0 No → **Find another survey respondent and go back to Q02.**
◯ 1 Yes

◯ 5 Household member not working on the holding → **Find another survey respondent and go back to Q02.**
◯ 6 Other (specify ______) → **Find another survey respondent and go back to Q02.**

TEXT TO READ:
This survey collects information about horticultural holdings in the whole country and whatever the holdings' sizes. Questions will focus mainly on areas used and horticultural crops production. To analyse this activity in its context, complementary questions will be asked about other crops produced, animals raised, meat, milk and eggs produced, persons working on the holding, other activities of the holding, income, expenditures and shocks. The reference period for the survey is the last complete agricultural year (country-specific start date and end date). (Note: if the survey is being conducted near the end of the current agricultural year, this will be the reference period).

Q03e Respondent's address (street) . . . . . □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
Q03f Respondent's address (village, town) . . □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

Q03g Telephone number (cell phone) . . . . □□□□□□□□□□□
Q03g Telephone number (landline) . . . . . □□□□□□□□□□□

Q03i E-mail address . . . . □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

Q04a. Is the holding currently growing any crops or fruits, or raising animals, or did it do so during the reference period?
◯ 0 No → **Go to Q05.**
◯ 1 Yes → **Go to Q04b.**

Q04b. Is the holding currently growing any horticultural crops, or did it do so during the reference period?

- ◯ 0 No → **Go to Section 4. Explain in the comments why the horticultural production had ceased and stop the questionnaire.**
- ◯ 1 Yes → **Go to Q07.**

Q05. Has the holding ceased its activity without any transfer of its means of production?

- ◯ 0 No → **Go to Q07.**
- ◯ 1 Yes

Q06. Will the holding resume its activity?

- ◯ 0 No → **Go to Section 4. Explain in the comments box how you tried to find the holding and stop the questionnaire.**
- ◯ 1 Yes

Q07. Are there any changes concerning the information known about the holding?

- ◯ 0 No → **Go to Part 1.2 Q10**
- ◯ 1 Yes

Q08. What changes occurred?

(Fill in one circle only)

- ◯ 1 Change of legal status → **Go to Part 1.2 Q10**
- ◯ 2 Transfer of all means of production to another single holding → **Go to Part 1.2 Q10**
- ◯ 3 Transfer of all means of production to several holdings → **Go to Q09a.**
- ◯ 4 Fusion or merger with one or several other holdings → **Go to Q09b.**
- ◯ 5 No change in legal status or address but some information must be corrected → **Go to Part 1.2 Q10**

Q09a. How many new holdings were created with the means of production of the original holding? . . . . ☐☐☐

Go to Part 1.2 Q10 *to record answers of the holding having the same address as the previous one or, if not possible, the holding having acquired the main part of the means of production.*

Q09b. How many other holdings merged with the original holding? . . . . ☐☐☐

Go to Part 1.2 Q10 *to record answers of the holding having the same address as the previous one or, if not possible, the holding having acquired the main part of the means of production.*

PART 1.2: IDENTIFICATION OF THE HOLDING

Q10. What is the legal status of the Holder?

(Fill in one circle only)
- ◯ 1 Civil person/natural person
- ◯ 2 Group of civil persons/natural persons
- ◯ 3 Legal person

Q11. What is the legal status of the holding?

(Fill in one circle only)
- ◯ 1 (Country-specific response option)
- ◯ 2 (Country-specific response option)
- ◯ 3 (Country-specific response option)

If Q10 = 1 or 2 GO TO Q12, otherwise GO TO Q14

Q12. Answer the following questions about the Holder/Co-holders.

Q12a First name . . .

Q12b Surname . . .

Q12c Sex
- ◯ 1 Male
- ◯ 2 Female

Q12d PERSONAL ID of the Holder . . . . .

Q13. Address of the Holder

Q13a. Region . . . . .

Q13b. District . . . . .

Q13c. Village or town name .

GO to Q15

Q14. What is the legal name of the holding? . . . .

Q15. Enumeration area of the holding . . .

Q16. Holding Serial Number . . . . . .

Q17. Address of the holding
- ◯ 1 Same as the address of the Holder → Go to Q18.
- ◯ 2 Different from the address of the Holder

Q17a Address (street) . . . . . .

Q17b Village, town . . . .

Q17c. Region . . . .

Q17d. District . . . .

Q17e. Main telephone number of the holding .

Q18. What is the main location type of the address reported above?

(Fill in one circle only)
- ◯ 1 Household dwelling (for HH sector) and farm, including dwelling and agricultural buildings
- ◯ 2 Main agricultural building
- ◯ 3 Main agricultural parcel

Q19. What are the GPS coordinates corresponding to the address of the holding?

Q21a Latitude . . . . ° .

Q21b Longitude . . . . ° .

Q20. What is the official identification number of the holding in the national business register? . . . .

Q21. What are the other administrative identification numbers of the holding?

Q21a Livestock . . . .

Q21b Wine production . . . .

Q21c Organic production . . . .

Q21d Other (specify) . . . .

Q22. What is the identification number of the holding from the last agricultural census? (can be prefilled)

PART 1.3: AGRICULTURAL ACTIVITY

Q23. **Does the holding record its agricultural activity or finances on registers or logbooks?**

(Fill in one circle only)
- ◯ 1 No, never → **Go to Q25.**
- ◯ 2 Yes, only occasionally or partially → **Go to Q25.**
- ◯ 3 Yes, systematically

Q24. What information is systematically registered?

(Fill in all that apply)
- ◯ 1 Area cultivated/harvested
- ◯ 2 Crop production
- ◯ 3 Livestock production
- ◯ 4 Unit prices, amounts sold and total sales by product
- ◯ 5 Input quantities used (seeds, fertilizers, plant protection products, etc.)
- ◯ 6 Detailed quantities and prices of inputs bought
- ◯ 7 Workers' time
- ◯ 8 Workers' payment
- ◯ 9 Other (specify ______)

Q25. What is the tenure of the agricultural land used by the holding during the reference period?

Q25a. What was the total area of the holding?

	Area	Unit of measure	Conversion factor to a standard unit	Area calculated in standard unit
a. Area used for crop productions	.			.
b. Kitchen gardens	.			.
c. Backyards	.			.
d. Farm buildings and farmyards	.			.
e. Forest and other wooded land	.			.
f. Aquaculture on the holding	.			.
g. Other land (unutilized, rocks, wetlands, etc.)	.			.
h. **Control total holding area (total of options a to g)**	.			.
i. **Control total holding area used for crop production (total of options a and b)**				.

Q25b. Of the total area used for crop production (Q25a. Option i), how much is:

	Area
a. **Owned with** written documentation (such as title deeds, wills, purchase agreements)	.
b. **Owned without** written documentation	.
c. **Rented**-in, leased or sharecropped **with** written agreement	.
d. **Rented**-in, leased or sharecropped **without** written agreement	.
e. State or communal land used **with** written agreement (certified use rights)	.
f. State or communal land used **without** written agreement (uncertified use rights)	.
g. Occupied/squatted without any permission	.
Control Total land (total of options a to g)	.

Q26. From an economic perspective, what is the holding's main agricultural focus for the reference period?
*** Answer based on the economic value of your activities, not the time spent on activities.**

(Fill in one circle only)
- ◯ 1 Mainly crop production
- ◯ 2 Mainly livestock production → **Go to Q28.**
- ◯ 3 A mix of crop and livestock production → **Go to Q29.**

Q27. From an economic perspective, what is the main cropping activity?
*** The main cropping activity is the one with the highest economic value.**

(Fill in one circle only)
- ◯ 1 Production of annual field crops (cereals, oilseeds, protein crops, root crops, tobacco, cotton, etc.) → **Go to Q29.**
- ◯ 2 Production of vegetables, mushrooms, flowers, ornamental plants, etc. → **Go to Q29.**
- ◯ 3 Production of grapes for wine → **Go to Q29.**
- ◯ 4 Production of fruits → **Go to Q29.**
- ◯ 5 Production of other perennial crops (cacao, coffee, etc.) → **Go to Q29.**
- ◯ 6 Mixed cropping (no real prevalence of a specific crop activity) → **Go to Q29.**

Q28. From an economic perspective, what is the main livestock activity?
*** The main livestock activity is the one with the highest economic value.**

(Fill in one circle only)
- ◯ 1 Raising ruminant livestock for meat (cattle, sheep, goats, etc.)
- ◯ 2 Raising non-ruminant livestock for meat (pigs, poultry, etc.)
- ◯ 3 Production of eggs
- ◯ 4 Production of milk
- ◯ 5 Mixed livestock (no real prevalence of a specific livestock activity)

Q29. What is the main intended destination of your agricultural production?

(Fill in one circle only)
- ◯ 1 Producing primarily for sale (selling 90% or more)
- ◯ 2 Producing mainly for sale, with some own consumption (selling more than 50% and up to 90%)
- ◯ 3 Producing mainly for own consumption, with some sales (selling more than 10% and up to 50%)
- ◯ 4 Producing primarily for own consumption (selling 10% or less)

Comments on SECTION 1:

SECTION 2: CHARACTERISTICS OF THE HOLDERS AND MANAGERS

CASE 1: THE HOLDER IS A CIVIL/NATURAL PERSON (SECTION 1, PART 1.2, Q10=1)

Q01. Provide the following information on the Holder.

Q01a First name . . . . [ ]
Q01b Surname . . . . [ ]

Q01c Contact number (preferably cell phone) . . . [ ]

Q01d Sex
- ◯ 1 Male
- ◯ 2 Female

Q01e Age in completed years . . . . [ ]

Q01f Nationality

(Fill in one circle only)
- ◯ 1 Local country
- ◯ 2 Neighbouring country
- ◯ 3 Other

Q01g Indigenous group

(Fill in one circle only)
- ◯ 1 (Country-specific response option)
- ◯ 2 (Country-specific response option)
- ◯ 3 (Country-specific response option)
- ◯ 4 None of the above

Q01h Highest level of education completed

(Fill in one circle only)
- ◯ 1 None
- ◯ 2 Less than primary
- ◯ 3 Primary
- ◯ 4 Lower secondary
- ◯ 5 Upper secondary
- ◯ 6 Tertiary/post-secondary

Q01i Share of working time spent working on the holding

(Fill in one circle only)
- ◯ 1 Less than half (< 40 %)
- ◯ 2 About half (40%-59%)
- ◯ 3 Most/almost all (60%-99%)
- ◯ 4 All (100%)

Q01j Does the Holder have another gainful activity outside of the holding?
- ◯ 0 No
- ◯ 1 Yes

Q01k Is the Holder also the Manager?
- ◯ 0 No → **Go to Q02.**
- ◯ 1 Yes → **Go to SECTION 3.**

Q02. Provide the following information on the Manager.

Q02a First name . . . . [ ]
Q02b Surname . . . . [ ]

Q02c Contact number (preferably cell phone) . . . [ ]

Q02d Sex
- ◯ 1 Male
- ◯ 2 Female

Q02e Age in completed years . . . . [ ]

Q02f Link with the Holder

(Fill in one circle only)
- ◯ 1 Wife/husband or consensual union partner
- ◯ 2 Other member of the household
- ◯ 3 External

Q02g Nationality

(Fill in one circle only)
- ◯ 1 Local country
- ◯ 2 Neighbouring country
- ◯ 3 Other

Q02h Indigenous group
(Fill in one circle only)
- ◯ 1 (Country-specific response option)
- ◯ 2 (Country-specific response option)
- ◯ 3 (Country-specific response option)
- ◯ 4 None of the above

Q02i Highest level of education completed

(Fill in one circle only)
◯ 1 None
◯ 2 Less than primary
◯ 3 Primary
◯ 4 Lower secondary
◯ 5 Upper secondary
◯ 6 Tertiary/post-secondary

Q02j Share of working time spent working on the holding

(Fill in one circle only)
◯ 1 Less than half (< 40 %)
◯ 2 About half (40%-59%)
◯ 3 Most/almost all (60%-99%)
◯ 4 All (100%)

Q02k Does the Manager have another gainful activity outside of the holding?
◯ 0 No
◯ 1 Yes

CASE 2: THE HOLDER IS A GROUP OF CIVIL/NATURAL PERSONS (SECTION 1, PART 1.2, Q10=2)

Q03. What is the number of civil/natural persons who are members of the Holder group? . . . ☐☐

REPEAT Q04. FOR ALL CO-HOLDERS (number reported in Q03).

Q04. Provide the following information for Co-Holder 1.

Q04a First name . . . .
Q04b Surname . . . .

Q04c Contact number (preferably cell phone) . . .

Q04d Sex
◯ 1 Male
◯ 2 Female

Q04e Age in completed years . . . . ☐☐☐

Q04f Nationality

(Fill in one circle only)
◯ 1 Local country
◯ 2 Neighbouring country
◯ 3 Other

Q04g Indigenous group

(Fill in one circle only)
◯ 1 (Country-specific response option)
◯ 2 (Country-specific response option)
◯ 3 (Country-specific response option)
◯ 4 None of the above

Q04h Highest level of education completed

(Fill in one circle only)
◯ 1 None
◯ 2 Less than primary
◯ 3 Primary
◯ 4 Lower secondary
◯ 5 Upper secondary
◯ 6 Tertiary/post-secondary

Q04i Share of working time spent working on the holding

(Fill in one circle only)
◯ 1 Less than half (< 40 %)
◯ 2 About half (40%-59%)
◯ 3 Most/almost all (60%-99%)
◯ 4 All (100%)

Q04j Does the Co-Holder 1 have another gainful activity outside of the holding?
◯ 0 No
◯ 1 Yes

Q04k Is the Co-Holder 1 also the Manager?
◯ 0 No
◯ 1 Yes

Q05. If there is no Manager among the Co-Holders, provide the following information on the Manager.

Q05a First name . . . .
Q05b Surname . . . .

Q05c Contact number (preferably cell phone) . . .

Q05d Sex
- ◯ 1 Male
- ◯ 2 Female

Q05e Age in completed years . . . . [| |]

Q05f Link with one of the Holders

(Fill in one circle only)
- ◯ 1 Wife/husband or consensual union partner
- ◯ 2 Other member of the household
- ◯ 3 External

Q05g Nationality

(Fill in one circle only)
- ◯ 1 Local country
- ◯ 2 Neighbouring country
- ◯ 3 Other

Q05h Indigenous group

(Fill in one circle only)
- ◯ 1 (Country-specific response option)
- ◯ 2 (Country-specific response option)
- ◯ 3 (Country-specific response option)
- ◯ 4 None of the above

Q05i Highest level of education completed

(Fill in one circle only)
- ◯ 1 None
- ◯ 2 Less than primary
- ◯ 3 Primary
- ◯ 4 Lower secondary
- ◯ 5 Upper secondary
- ◯ 6 Tertiary/post-secondary

Q05j Share of working time spent working on the holding

(Fill in one circle only)
- ◯ 1 Less than half (< 40 %)
- ◯ 2 About half (40%-59%)
- ◯ 3 Most/almost all (60%-99%)
- ◯ 4 All (100%)

Q05k Does the Manager have another gainful activity outside of the holding?
- ◯ 0 No
- ◯ 1 Yes

CASE 3: THE HOLDER IS A LEGAL PERSON (SECTION 1, PART 1.2, Q10=3)

Q06. What is the number of civil/natural persons participating in the capital of the company? . . . . [|]

Q07. What is the number of legal persons participating in the capital of the company? . . . . [|]

Q08. How many Managers are associated with the holding? . . . . [|]

REPEAT Q09. for each MANAGER (number reported in Q08) .

Q09. Provide the following information for each Manager.

Q09a First name . . . . [| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |]

Q09b Surname . . . . [| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |]

Q09c Contact number (preferably cell phone) . . . [| | | | | | | | | | | | |]

Q09d Sex
- ◯ 1 Male
- ◯ 2 Female

Q09e Age in completed years . . . . [| |]

Q09f Nationality

(Fill in one circle only)
- ◯ 1 Local country
- ◯ 2 Neighbouring country
- ◯ 3 Other

Q09g Indigenous group

(Fill in one circle only)
- ◯ 1 (Country-specific response option)
- ◯ 2 (Country-specific response option)
- ◯ 3 (Country-specific response option)
- ◯ 4 None of the above

Q09h Highest level of education completed

(Fill in one circle only)

◯ 1 None
◯ 2 Less than primary
◯ 3 Primary
◯ 4 Lower secondary
◯ 5 Upper secondary
◯ 6 Tertiary/post-secondary

Q09i Share of working time spent working on the holding

(Fill in one circle only)

◯ 1 Less than half (< 40 %)
◯ 2 About half (40%-59%)
◯ 3 Most/almost all (60%-99%)
◯ 4 All (100%)

Q09j Does the Manager have another gainful activity outside of the holding?

◯ 0 No
◯ 1 Yes

Comments on SECTION 2:

SECTION 3: HORTICULTURAL CROPS PRODUCTION DURING THE REFERENCE PERIOD DD/MM/YYYY to DD/MM/YYYY

PART 3.1: VEGETABLES AND ORNAMENTAL PLANTS PRODUCTION

Q00. Did the holding grow vegetables during the reference period, whatever the production or destination?

- ◯ 0 No → Go to Q00a.
- ◯ 1 Yes → Go to Q01.

Q00a. Do you plan to introduce vegetables production in the upcoming period?

- ◯ 0 No → Go to section 4.
- ◯ 1 Yes → Go to Q01.

Q01. What is the area of the holding used for vegetables and/or ornamental plants production?

	Area	Unit of measure	Conversion factor to a standard unit	Area calculated in standard unit	Number of beds
Q01a Vegetables and/or ornamental plants under greenhouses or high shelters					
Q01b Vegetables and/or ornamental plants under hydroponic cropping					
Q01c Vegetables and/or ornamental plants outdoors or under low shelters					
Q01d Vegetables and/or ornamental plants in kitchen garden					
Total agricultural area utilized for vegetables *(calculated)*					

Q02. What vegetables and/or ornamental plants were produced on the holding during the reference period?

*** Report all vegetables and/or ornamental plants grown, regardless of the quantity harvested (even zero) or the destination.**
*** Report associated crops grown on the same parcel.**
*** See vegetables and ornamental plants list.**

	Vegetable and/or ornamental plants name	Vegetable code
[VEGETABLE 1]		
[VEGETABLE 2]		
[VEGETABLE 3]		
[VEGETABLE 4]		
[VEGETABLE 5]		
[VEGETABLE 6]		
[VEGETABLE 7]		
...	...	...

During the first visit at the beginning of the reference period, the list is initialized.
During the visit of the end of reference period, the list is updated following the plants actually grown up.

PART 3.2: HARVEST AND USE OF VEGETABLES AND ORNAMENTAL PLANTS PRODUCED DURING THE REFERENCE PERIOD

*** The following questions will be answered for each vegetable and/or ornamental plant grown, regardless of the quantity harvested (even zero) or the destination.**

Q03. What is the total area sown or planted for [VEGETABLE]?
(For example, if a bed of 200 m² has been sown twice, the area will be 400 m² here)

Area	Unit of measure	Conversion factor to a standard unit	Area calculated in standard unit	Number of beds

Q03a Among this area, what part was:

1 Under greenhouses or high shelters ___ %
2 Under hydroponic technique ___ %
3 Under low shelters (at least some times) ___ %
4 Always in open field ___ %

5 Irrigated ___ %
6 Drained ___ %

7 Under official organic farming ___ %

Q04. Were fertilizers used on [VEGETABLE]?

- ◯ 0 No
- ◯ 1 Yes

Q04a "Yes", what type of fertilizers?
(Fill in all that apply)

- ◯ 1 Mineral
- ◯ 2 Organo-mineral
- ◯ 3 Compost
- ◯ 4 Mulch
- ◯ 5 Organic
- ◯ 6 Animal effluents
- ◯ 7 Other (specify ______)

Q05. Were plant protection products (PPP) used on [VEGETABLE]?

- ◯ 0 No
- ◯ 1 Yes

Q05a "Yes", what type of PPPs?
(Fill in all that apply)

- ◯ 1 Insecticide
- ◯ 2 Herbicide
- ◯ 3 Fungicide
- ◯ 4 Rodenticide
- ◯ 5 Other (specify ______)

	Quantity	Unit of measure	Conversion factor to a standard unit	Quantity calculated in standard unit
Q06. What was the quantity of [VEGETABLE] harvested?				

Q06a How was the yield of [VEGETABLE] compared to the same harvest of the previous year?

(Fill in one circle only)
- ◯ 1 Similar
- ◯ 2 Greater
- ◯ 3 Lower

Q07. What were the different uses of harvested [VEGETABLE]?

	Percentage	OR	Quantity	Unit of measure	Conversion factor to a standard unit	Quantity calculated in standard unit
Q07a Self-consumption	%					
Q07b Used as pay or wages for labour	%					
Q07c Given to service or input providers	%					
Q07d Used for on farm processing	%					
Q07e Direct sale to consumers	%					
Q07f Sold to industry, wholesaler or reseller	%					
Q07g Losses after harvest	%					
Total (calculated)	1 0 0 %					

Q08. If some part of [VEGETABLE] has been sold directly to consumers, what was the average price?

Total amount		Average unit price
	OR	

Q09. If some part of [VEGETABLE] has been sold to industry, wholesaler or reseller, what was the average price?

Total amount		Average unit price
	OR	

Comments on SECTION 3:

SECTION 4: END OF THE SURVEY

PART 4.1 SURVEY TIMING - TO BE COMPLETED BY THE SURVEYOR

End time of the survey: ☐☐ hour ☐☐ minutes

Survey duration ☐☐☐ (calculated)

PART 4.2 RESPONDENT OPINION OF SURVEY BURDEN

Q01. What is your overall judgement on the difficulty of this survey?

(Fill in one circle only)
- ◯ 1 acceptable
- ◯ 2 too difficult

Q02. What is your overall judgement on the length of this survey?

(Fill in one circle only)
- ◯ 1 acceptable
- ◯ 2 too long

End of survey, thank you.

General comments on the survey:

Annex 12

Crop production section of the AGRIS core module questionnaire

Global Strategy to improve Agricultural and Rural Statistics

AGRIS QUESTIONNAIRE
CORE MODULE

Version 1.1, September 2017

SECTION 3: CROP PRODUCTION DURING THE REFERENCE PERIOD DD/MM/YYYY to DD/MM/YYYY

PART 3.1: CROP PRODUCTION AND DESTINATIONS

Q00. Did the holding grow crops during the reference period, whatever the production or destination?
- ◯ 0 No → **Go to Q00a.**
- ◯ 1 Yes → **Go to Q01.**

Q00a. What was the area of the holding used for other pruposes than crop production?

	Area	Unit of measure	Conversion factor to a standard unit	Area calculated in standard unit
Kitchen gardens and backyards				
Farm buildings and farmyards				
Forest and other wooded land				
Aquaculture on the holding				
Other land (unutilised, rocks, wetlands, etc.)				

Q00b. Do you plan to introduce crops in the upcomming period?
- ◯ 0 No → **Go to section 4.**
- ◯ 1 Yes → **Go to Q17a, Q17b and Q17c.**

Q01. Would you be confident in providing an estimate of the area of your holding?
- ◯ 0 No → **Please give the best estimations you can to the following questions.**
- ◯ 1 Yes

Q02. How many parcels did you use for agricultural production (for crops and livestock) during the reference period? . . . [| |]

Q03. What crops were produced on the holding during the reference period?

*** Report all crops grown, regardless of the quantity harvested (even zero).**
*** Report associated crops grown on the same parcel.**
*** See crop list.**

	Crop name	Crop code
[CROP 1]		
[CROP 2]		
[CROP 3]		
[CROP 4]		
[CROP 5]		
[CROP 6]		
[CROP 7]		

Q04. Answer the following questions about [CROP].

Q04a Were fertilizers used on [CROP]?
- ◯ 0 No
- ◯ 1 Yes

Q04b Were plant protection products used on [CROP]?
- ◯ 0 No
- ◯ 1 Yes

Q04c Did the holding have a stock of [CROP] stored just before the last harvest?
*** Include only stock owned by the holding.**
*** Include stock stored on the holding and off of the holding.**
- ◯ 0 No → **Go to Q04f.**
- ◯ 1 Yes

	Quantity	Unit of measure	Conversion factor to a standard unit	Quantity calculated in standard unit
Q04d How much [CROP] was stored on the holding				
Q04e How much [CROP] was stored at a location off of the holding				

Q04f How many harvests of [CROP] were there in the reference period?

(Fill in one circle only)
- ◯ 1 Continuous harvest → **Go to Q05.**
- ◯ 2 One harvest → **Go to Q06.**
- ◯ 3 Two harvests → **Go to Q06.**
- ◯ 4 Three harvests → **Go to Q06.**
- ◯ 5 Four harvests → **Go to Q06.**
- ◯ 6 No harvest → **Go to NEXT CROP.**

Q05. Answer the following questions on continuous harvest [CROP].

*** The reference period for questions on continuous harvest crops is the last six months.**

	Area	Unit of measure	Conversion factor to a standard unit	Area calculated in standard unit
Q05a What area of [CROP] was planted in the last six months?				

Q05b Was [CROP] irrigated during the last six months?
- ◯ 0 No
- ◯ 1 Yes

	Quantity	Unit of measure	Conversion factor to a standard unit	Quantity calculated in standard unit
Q05c What was the quantity of [CROP] harvested in the last six months?				

Q05d How was the production of [CROP] compared to the previous six months?

(Fill in one circle only)
- ◯ 1 Similar
- ◯ 2 Greater
- ◯ 3 Lower

Q05e Was [CROP] cultivated together with other crops (at the same time in the same parcel)?
- ◯ 0 No → **Go to Q10.**
- ◯ 1 Yes, for all of the crop → **Go to Q10.**
- ◯ 2 Yes, for a part of the crop → **Go to Q10.**

Q06. Answer the following questions about the most recent harvest (Harvest 1) of [CROP].

*** Include crops harvested at least once during the reference period.**

*** Exclude continuous harvest crops.**

Q06a When did the last harvest start for [CROP]? YYYY / MM / DD

Q06b How many days did the harvest of [CROP] last? days

Q06c Was [CROP] irrigated during this harvest season?
- ◯ 0 No
- ◯ 1 Yes

	Area	Unit of measure (see codes)	Conversion factor to a standard unit	Area calculated in standard unit
Q06d What area of [CROP] was planted?				
Q06e What area of [CROP] was harvested?				

	Quantity	Unit of measure	Conversion factor to a standard unit	Quantity calculated in standard unit
Q06f What was the quantity of [CROP] harvested?				

Q06g How was the yield of [CROP] compared to the same harvest of the previous year?

(Fill in one circle only)
- ◯ 1 Similar
- ◯ 2 Greater
- ◯ 3 Lower

Q06g Was [CROP] cultivated together with other crops (at the same time in the same parcel)?
- ◯ 0 No
- ◯ 1 Yes, for all of the crop
- ◯ 2 Yes, for a part of the crop

FOR CROPS THAT HAD ONE HARVEST IN THE LAST REFERENCE PERIOD (Q04f=2) → **Go to Q10.**

FOR CROPS THAT HAD MORE THAN ONE HARVEST IN THE LAST REFERENCE PERIOD (Q04f=3, 4, 5) → **Go to Q07.**

Q07. Answer the following questions about the harvest before the most recent harvest (penultimate) (Harvest 2) of [CROP].

*** Include crops harvested at least twice during the reference period .**

*** Exclude continuous harvest crops.**

Q07a When did the penultimate harvest (Harvest 2) start for [CROP]? YYYY / MM / DD

Q07b Was [CROP] irrigated during this harvest season?
- ◯ 0 No
- ◯ 1 Yes

	Area	Unit of measure (see codes)	Conversion factor to a standard unit	Area calculated in standard unit
Q07c What area of [CROP] was planted?				
Q07d What area of [CROP] was harvested?				

	Quantity	Unit of measure	Conversion factor to a standard unit	Quantity calculated in standard unit
Q07e What was the quantity of [CROP] harvested?				

Q07f Was [CROP] cultivated together with other crops (at the same time in the same parcel)?
- ◯ 0 No
- ◯ 1 Yes, for all of the crop
- ◯ 2 Yes, for a part of the crop

FOR CROPS THAT HAD TWO HARVESTS IN THE LAST REFERENCE PERIOD (Q04f=3) → Go to Q10.
FOR CROPS THAT HAD MORE THAN TWO HARVESTS IN THE LAST REFERENCE PERIOD (Q04f=4, 5) → Go to Q08.

Q08. Answer the following questions about the antepenultimate harvest (Harvest 3) of [CROP].
*** Include crops harvested at least three times during the reference period .**
*** Exclude continuous harvest crops.**

Q08a When did the antepenultimate harvest (Harvest 3) start for [CROP]? . . . . . Y Y Y Y / M M / D D

Q08b Was [CROP] irrigated during this harvest season?
- ◯ 0 No
- ◯ 1 Yes

	Area	Unit of measure (see codes)	Conversion factor to a standard unit	Area calculated in standard unit
Q08c What area of [CROP] was planted?				
Q08d What area of [CROP] was harvested?				

	Quantity	Unit of measure	Conversion factor to a standard unit	Quantity calculated in standard unit
Q08e What was the quantity of [CROP] harvested?				

Q08f Was [CROP] cultivated together with other crops (at the same time in the same parcel)?
- ◯ 0 No
- ◯ 1 Yes, for all of the crop
- ◯ 2 Yes, for a part of the crop

FOR CROPS THAT HAD THREE HARVESTS IN THE LAST REFERENCE PERIOD (Q04f=4) → Go to Q10.
FOR CROPS THAT HAD FOUR HARVESTS IN THE LAST REFERENCE PERIOD (Q04f=5) → Go to Q09.

Q09. Answer the following questions about the oldest harvest (Harvest 4) of [CROP] during the reference period.
*** Include crops harvested at least four times during the reference period .**
*** Exclude continuous harvest crops.**

Q09a When did the oldest harvest (Harvest 4) start for [CROP]? . . . . . Y Y Y Y / M M / D D

Q09b Was [CROP] irrigated during this harvest season?
- ◯ 0 No
- ◯ 1 Yes

	Area	Unit of measure (see codes)	Conversion factor to a standard unit	Area calculated in standard unit
Q09c What area of [CROP] was planted?				
Q09d What area of [CROP] was harvested?				

	Quantity	Unit of measure	Conversion factor to a standard unit	Quantity calculated in standard unit
Q09e What was the quantity of [CROP] harvested?				

Q09f Was [CROP] cultivated together with other crops (at the same time in the same parcel)?
- ◯ 0 No
- ◯ 1 Yes, for all of the crop
- ◯ 2 Yes, for a part of the crop

Q10. Answer the following questions about the destinations of the holding's production from all harvests during the reference period.
*** Use the same unit of measure that was reported for quantities in previous questions.**

	Quantity	Unit price of the last sale	Unit used to describe the price
Q10a What was the quantity of [CROP] for own use? *** Include farm use for seeds or animal feed. * Include household and other family use.**			
Q10b What was the quantity of [CROP] sold?			
Q10c What was the quantity of [CROP] used as pay for labour as wages?			
Q10d What was the quantity of [CROP] given to service or input providers for pay (land, seeds, plant protection products, fertilizers, etc.)?			

PART 3.2: AREA UTILIZED

Q11. What was the area used for the following agricultural purposes?

*** Refer to the last harvest for Q11a to Q11e.**

	Agricultural area utilized	Area	Unit of measure	Conversion factor to a standard unit	Area calculated in standard unit
Q11a	Temporary crops under greenhouses or high shelters	.			.
Q11b	Temporary crops outdoors or under low shelters	.			.
Q11c	Temporary fallow	.			.
Q11d	Temporary meadows and pastures	.			.
Q11e	Kitchen gardens and backyards	.			.
Q11f	Permanent crops under greenhouses or high shelters	.			.
Q11g	Permanent crops outdoors or under low shelters	.			.
Q11h	Permanent meadows and pastures	.			.
	Total agricultural area utilized (AAU) *(calculated)*	.			.

Q11i Do you confirm that the calculated area corresponds to the holding's total agricultural area utilized?
- ◯ 0 No → **Ask Q11 again.**
- ◯ 1 Yes

		Area	Unit of measure	Conversion factor to a standard unit	Area calculated in standard unit
Q11j	Area equipped for irrigation in working order	.			.
Q11k	Area equipped with soil drainage in working order	.			.
Q11l	Area in organic farming (part of the AAU calculated above)	.			.

Q12. If there is area equipped for irrigation in working order, did you irrigate during the reference period?
- ◯ 0 No
- ◯ 1 Yes

Q13. Was there land used for the following purposes?

Q13a Farm buildings and farmyards
- ◯ 0 No
- ◯ 1 Yes

Area [.]

Q13b Forest and other wooded land
- ◯ 0 No
- ◯ 1 Yes

Area [.]

Q13c Aquaculture on the holding (area not included in Q11)
- ◯ 0 No
- ◯ 1 Yes

Area [.]

Q13d Other land (unutilized, rocks, wetlands, etc.)
- ◯ 0 No
- ◯ 1 Yes

Area [.]

PART 3.3: CROP PRODUCTION MODES

Q14. Answer the following questions about the type of seeds used for [CROP].

*** Include all temporary crops.**

*** Exclude perennial crops and permanent pastures.**

Q14a How many varieties of [CROP] were used?

(Fill in one circle only)

◯ 1 One variety

◯ 2 More than one variety

	Percent	
Q14b What share of the [CROP] seed consisted in certified modern varieties?		%
Q14c What share of the [CROP] seed consisted in uncertified varieties?		%
Total	1 0 0	%

Q14d If certified varieties of [CROP] were used, were any of them Genetically Modified Organisms (GMO)?

◯ 0 No

◯ 1 Yes

THE ANSWER TO THIS QUESTION WILL DETERMINE IF Q15a ETC. WILL BE ASKED FOR CROP1 AND SUBSEQUENT CROPS.

Q15. Does the holding have any production and/or marketing contracts for any crops?

◯ 0 No **→ Go to Q16. DO NOT INCLUDE Q15 OR ITS SUBQUESTIONS FOR ANY SUBSEQUENT CROPS**

◯ 1 Yes **→ Go to Q15a.**

For each of crops identified in Q06, answer the following questions:

Q15a Does the holding have a production and/or marketing contract for [CROP]?

◯ 0 No **→ Go to next crop, after the last crop grown go to Q16.**

◯ 1 Yes **→ Go to Q15b.**

Q15b Does the holding have a production contract for [CROP]?

◯ 0 No **→ Go to Q15d.**

◯ 1 Yes **→ Go to Q15c.**

Q15c Does the production contract cover 100% of the [CROP] grown by the holding (exclusive contract)?

◯ 0 No

◯ 1 Yes

Q15d Does the holding have a marketing contract for [CROP]?

◯ 0 No **→ Go to next crop, after the last crop grown go to Q16.**

◯ 1 Yes **→ Go to Q15e.**

Q15e Does the marketing contract cover 100% of the [CROP] grown by the holding (exclusive contract)?

◯ 0 No

◯ 1 Yes

PART 3.4: INTENTIONS FOR CROP PRODUCTION FOR THE 12 MONTHS AFTER THE REFERENCE PERIOD

Q16. For each of the crops identified in Q06 answer the following questions about planting intentions for [CROP].

Q16a What area do you plan to dedicate to [CROP] in the upcoming period?

(Fill in one circle only)
- ◯ 1 Similar → **Go to Q17.**
- ◯ 2 Greater
- ◯ 3 Lower
- ◯ 4 None

16b What is the main reason for changes in the intended area of [CROP]?

(Fill in one circle only)
- ◯ 1 Crop rotation
- ◯ 2 Technical
- ◯ 3 Economic
- ◯ 4 Other (specify ____________)

Q17. Do you plan to introduce OTHER crops in the upcoming period (crops not identified in Q06)?
- ◯ 0 No → **Go to SECTION 4.**
- ◯ 1 Yes

Q17a What other crops do you plan to introduce in the upcoming period?

	Crop name	Crop code
[CROP 1]		
[CROP 2]		
[CROP 3]		
[CROP 4]		

	Area	Unit of measure (see codes)	Conversion factor to a standard unit	Area calculated in standard unit
Q17b What area of [CROP] do you plan to cultivate?				

Q17c What is the main reason for the planned introduction of [CROP]?

(Fill in one circle only)
- ◯ 1 Crop rotation
- ◯ 2 Technical
- ◯ 3 Economic
- ◯ 4 Other (specify ____________)

THE SERIES OF CROP-RELATED QUESTIONS WILL BE ASKED FOR EACH OF THE CROPS IDENTIFIED IN Q06. ONCE COMPLETE FOR ALL CROPS, PROCEED TO SECTION 4

Comments on SECTION 3:

Layout:

- Laura Monopoli

Cover photos:

- © Asian Development Bank
- © CIAT/Neil Palmer
- © FAO/Olivier Thuillier