



**Food and Agriculture Organization  
of the United Nations**

# **Methodology for Estimation of Crop Area and Crop Yield under Mixed and Continuous Cropping**

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

March 2017

**Methodology for Estimation  
of Crop Area and Crop Yield  
under Mixed and Continuous  
Cropping**

# Table of Contents

<b>Abstract</b> .....	<b>4</b>
<b>Acknowledgements</b> .....	<b>5</b>
<b>Executive summary</b> .....	<b>6</b>
<b>Acronyms and Abbreviations</b> .....	<b>8</b>
<b>1.</b>	<b>9</b>
<b>Introduction</b> .....	
1.1	
Background.....	9
1.2 Objectives.....	12
1.3 Proposed approach.....	12
<b>2. Methods of crop area apportioning under mixed cropping and intercropping....</b>	<b>13</b>
2.1 Area apportioning using an objective method.....	14
2.2 Area apportioning using a subjective method.....	17
<b>3. Area measurement methods and data collection</b> .....	<b>21</b>
3.1	
Introduction.....	21
3.2 Measurement methods used for crop area estimation.....	21
<b>4 Yield measurement methods and data collection</b> .....	<b>27</b>
4.1	27
Introduction.....	
4.2 Measurement methods for yield estimation.....	27
<b>5. Methodology for estimating crop area and crop yield under mixed and continuous cropping</b> .....	<b>36</b>
5.1	36
Introduction.....	
5.2 Sampling design used for crop area estimation in mixed cropping.....	36
<b>6. Results of three country field tests</b> .....	<b>58</b>
6.1	58
Introduction.....	
6.2 Results of the field tests.....	58
6.3 Results from world bank methodological experiments in Africa on area measurement.....	73
<b>7. Recommendations</b> .....	<b>75</b>
<b>Conclusions</b> .....	<b>77</b>
<b>References</b> .....	<b>78</b>
<b>Glossary</b> .....	<b>83</b>

# Abstract

On the basis of a gaps analysis, a methodology has been developed to estimate crop area and crop yield in mixed and continuous cropping scenarios. In this regard, several alternatives have been considered, depending upon the information available in the agricultural statistical system. The different methods for the area apportionment of a crop mixture's various component crops are explained, as are methods for crop area and yield measurement, along with their respective advantages and disadvantages. Situations in which particular methods are suitable are described. The questionnaires mentioned in this technical report have been designed for data collection on crop area and crop yield for the alternative methodologies developed. The importance of Computer-Assisted Personal Interviewing (CAPI) software for the efficient collection of survey data is emphasized. The methodology developed for crop area and yield estimation is demonstrated through a series of field tests conducted in district/study areas in Indonesia, Jamaica and Rwanda. The report ends with a number of Conclusions.

# Acknowledgements

This document is prepared by U.C. Sud, Tauqueer Ahmad, V.K. Gupta, Hukum Chandra, Prachi Misra Sahoo, Kaustav Aditya, Man Singh and Ankur Biswas of ICAR-Indian Agricultural Statistics Research Institute, New Delhi, India. The authors wish to express their gratitude and appreciation to the Food and Agriculture Organization of the United Nations (FAO) for awarding this study to the ICAR-Indian Agricultural Statistics Research Institute (ICAR-IASRI) in New Delhi. The authors also wish to thank Mr Naman Keita, Mr Christophe Duhamel and Mr Michael Austin Rahija of the Global Strategy for their constant support towards the efficient and timely completion of various activities pertaining to this study; Mr Alberto Zezza and Ms Sydney Gourlay (World Bank) for sharing the data relating to the study conducted by the World Bank team; and Ms Consuelo Señoret (Global Strategy) for her continuous administrative support, crucial for the completion of the project.

The authors also acknowledge with gratitude the support provided by Dr Harcharan Singh towards improving the quality of the technical report; the contribution of Mr D.P. Sharma, Chief Technical Officer (ICAR-IASRI) in implementing the CAPI software in the field-test countries; as well as the suggestions made by the peer reviewers and experts who attended the Expert Meeting held at FAO headquarters on April 15-17, 2015, which led to considerable improvements in the presentation of the technical report's contents.

# Executive Summary

In developing countries, agriculture tends to be the most important segment of the national economy. Agricultural statistics are necessary to provide information for monitoring trends and estimating the future prospects of agricultural commodity markets, and thus to assist in setting policies on aspects such as price support, imports and exports, and distribution. In particular, crop statistics (i.e. those on crop area, yield and production) play an important role in the planning and allocation of resources for the development of the agricultural sector. In developing and under-developed countries, the availability and quality of agricultural statistics has been declining; some countries even lack the capacity to produce a minimum set of data, as evidenced by the poor response rates to questionnaires formulated by the Food and Agriculture Organization of the United Nations (FAO; FAO, 2010). The Global Strategy to improve Agricultural and Rural Statistics (hereafter, Global Strategy) is a ground breaking effort aimed at strengthening agricultural statistics. At its Forty-first Session in February 2010, the United Nations Statistical Commission (UNSC) endorsed the Global Strategy's technical content and strategic direction, urging the rapid development of an Action Plan (hereafter, Global Action Plan) for its implementation. One of the issues identified under the Global Strategy's Research component is the estimation of crop area, yield and production in the context of mixed, repeated and continuous cropping. Accordingly, a study project entitled "Improving Methods for Estimating Crop Area, Yield and Production under Mixed, Repeated and Continuous Cropping" was awarded to the ICAR-Indian Agricultural Statistics Research Institute, New Delhi.

Under this study project, several technical reports are being produced. This technical report is the sixth in the series<sup>1</sup> and addresses the results of the gap analysis explored in Technical Report No. 2, *Gap Analysis on Improving Methods for Estimating Crop Area, Yield and Production under Mixed, Repeated and Continuous Cropping*, proposing an appropriate methodology for the estimation of crop area and crop yield under mixed and continuous cropping. In particular, in the domain estimation approach proposed, various crop mixtures are considered as domains. Further, an objective method is proposed for the apportioning of the crop mixture area into component crops. Methods are suggested for apportioning crop area in case of mixtures of annual and seasonal crops, annual and annual crops, annual and perennial crops, and perennial and perennial crops. Different measurement methods are proposed for determining

---

<sup>1</sup> For the complete list of published technical reports, see <http://gsars.org/en/tag/crops/>.

crop area and crop yield, and their respective advantages and disadvantages examined. A sample-survey-based approach for estimating crop area and crop yield is suggested. Estimators based on a double-sampling approach are proposed to estimate crop area and crop yield, combining subjective and objective methods of measuring these two aspects. Based on the prevalent agricultural statistical system, two approaches have been recommended for crop area and yield estimation for mixed and continuous cropping: (1) the area frame approach and (2) the household approach.

The methodology developed for crop area and crop yield estimation was field-tested in Indonesia, Rwanda and Jamaica. The results of these field tests are detailed in Technical Report V and results from similar studies performed by the World Bank are incorporated. Appropriate recommendations have been made based on the results of these studies.

The problem of continuous cropping could not be studied due to the short field testing period as well as operational difficulties like lack of enough trained enumerators, small sample sizes etc. encountered in the field. In the absence of data on physical observation, apportioning of crop mixture area into component crops was carried out using the method of seed rates, which may not be very objective method as information on seed rates is provided by the farmer. A limitation of the study is that small sample sizes were used for estimation of crop area and yield. Further, the data collection period was too short which created difficulties in implementation of field work. Also, the countries concerned could not collect data on many variables including gold standard for crop area estimation. Mixed cropping was not found in some of the districts/study areas. This highlights the need to replicate the study in future so that the problem of estimation of crop area and yield in mixed and continuous cropping can be studied properly with relevant details.

# Acronyms and Abbreviations

BPS	<i>Badan Pusat Statistik</i> (Indonesia)
CAPI	Computer-Assisted Personal Interviewing
CB	Census Block
CCE	Crop-Cutting Experiment
CCM	Crop Card Monitor
CV	Coefficient of Variation
EA	Enumeration Area
ED	Enumeration District
FAO	Food and Agriculture Organization of the United Nations
GPS	Global Positioning System
ICAR- IASRI	ICAR-Indian Agricultural Statistics Research Institute
LSMS	Living Standard Measurement Study
PPSWR	Probability Proportional to Size With Replacement
PSU	Primary Sampling Unit
PAPI	Paper-Assisted Personal Interviewing
SRSWOR	Simple Random Sampling Without Replacement
SSU	Secondary Sampling Unit
USU	Ultimate Sampling Unit

# Introduction

## 1.1. Background

Information on crop area, yield and production plays a vital role in planning and allocating resources for the development of the agricultural sector. Reliable and timely information on crop area, yield and production acts as a fundamental input to the planners and policymakers responsible for formulating efficient agricultural policies, and for making important decisions with respect to procurement, storage, public distribution, import, export and other related issues. The availability of crop area statistics is an essential requirement of the agricultural statistical system of any country, as it is a key variable in estimating crop production and crop yield. For the collection of crop area statistics, both subjective and objective methods are currently used around the world. The subjective methods, often used in developing countries, include the field reporting system, eye estimation, farmer interview and expert assessment. These methods suffer from certain limitations in terms of the reliability of the data on crop area. Although objective methods of measuring area – such as the polygon method – are expected to provide reliable estimates, they are costly and time-consuming. Further, under certain unusual and problematic situations (e.g. fields with irregular shapes and boundaries), it becomes difficult to measure area with subjective methods. In these cases, modern technologies such as Global Positioning Systems (GPS) have the potential to provide more accurate estimates of the crop area. However, the accuracy of the GPS method is known to be limited in certain conditions, such as when determining the crop area under cloudy conditions, in hilly regions where crops are grown on slopes, or in areas surrounded by trees.

As in case of crop area determination, both subjective and objective methods are currently adopted to collect yield statistics in various countries. The subjective methods of estimating crop yield include farmers' assessments, expert opinions and crop cards, while the objective methods include whole-plot harvesting and crop-cutting experiments.

The practice of sowing crops in mixture in a single parcel of land is prevalent in many countries, particularly where land holdings are small. The growing of crops in mixtures is a common practice because it protects farmers from adverse

weather conditions such as drought, flood, and pest and disease infestation. Further, it enables maximal utilization of the space, moisture and nutrients available in the field. Cultivators usually mix crops that cannot withstand a particular type of weather with another set of crops that thrive under those same conditions. In West Africa and the Sahel, it has been estimated that 60 to 80 per cent of parcels contain a mixture of crops. The methods employed for sowing such crops vary, not only from region to region but also from area to area – and even field to field – within the same region. The crops in the mixture are sown either individually in separate rows (intercropping) or mixed together. In the former case, the seeds of constituent crops are kept separate and a certain number of rows of one crop alternate with those of another. In the latter case, the seeds of two or more crops are mixed together before sowing and the mixed seeds are either sown in a row or broadcasted. Calculating the area for each crop in the crop mixture becomes more complex when the number of crops in mixture or in association increases, the proportions of different crops in mixture vary from field to field, the sowing/planting and harvesting time differ, and the crops' growing periods (the vegetative cycle lesser than 3 months to over one year) are of different lengths. The number of crops in mixture in the field may vary depending on the growing period of the constituent crops. Accordingly, it is not possible to estimate the crop area of the constituent crops of a mixture in a single visit. On the other hand, failure to take into account the constituent crops may result in grossly underestimated outputs. For example, in Niger, pulses are intercropped with cereals. Enumerating only one principal crop would result in capturing only 74 per cent of the output (Just, 1981).

An even distribution of rainfall allows farmers to sow and harvest crops throughout the year. The practice of successively sowing and harvesting the same or different crops on the same piece of land during the agricultural year is known as continuous cropping. This method of cropping is popular with farmers because it enables the cultivation of more than one crop on the same piece of land (thus without the need to acquire more land), and is considered to be more profitable than traditional cropping systems (FAO, 1982). Continuous cropping is practiced when it is lucrative to grow a single crop or when the demand for alternative crops is limited. Continuous cropping requires expertise in terms of crop management, farm implementation and equipment, marketing knowledge, etc. (Sheaffer & Moncada, 2011, pp. 340-341).

## 1.2. Objectives

As noted above, in view of the importance of estimating crop area, yield and production under mixed and continuous cropping, the Global Strategy has awarded the study project entitled “Improving Methods for Estimating Crop Area, Yield and Production under Mixed and Continuous Cropping” to ICAR-IASRI. The study project has the following objectives:

1. Critically review the literature pertaining to crop area and yield under mixed and continuous cropping;
2. Identify the gaps relating to the estimation of crop area and yield under mixed and continuous cropping;
3. Develop a standard statistical methodology for the estimation of the area and yield rate under mixed and continuous cropping;
4. Test the developed methodology in three field-testing countries in Asia-Pacific, Africa and the Latin America/Caribbean region (one country in each region);
5. Identify issues and challenges and provide suitable guidelines for the implementation of the developed methodology in developing countries.

## 1.3. Proposed approach

The proposed approach focuses on combining objective and subjective methods of measurement of crop area and yield or production with the optimal use of statistical procedures. Furthermore, the sample survey approach is suggested for estimating the area and yield of crops grown under mixed and continuous cropping. The methodology for crop area and yield estimation for mixed and continuous cropping must be developed under the following three scenarios:

1. Cadastral maps are available;
2. The area frame is available; and
3. Information on parcels is not available in records (i.e. the household approach).

Three different approaches are suggested to tackle each of the above scenarios.

With respect to the area frame and the household approach, the developed methodology was tested in three countries: Indonesia, Rwanda and Jamaica. Keeping the existing agricultural statistics systems of the countries in mind, the household approach was proposed in Indonesia and Jamaica, whereas the area frame approach was suggested for Rwanda. Thus, the presentation of

methodological developments here will be limited to the area frame and household approaches.

Two different approaches were adopted for primary data collection: (i) Paper-Assisted Personal Interviewing (PAPI); and (ii) Computer-Assisted Personal Interviewing (CAPI). A traditional PAPI-based approach was followed in all three countries, whereas a CAPI-based approach was also followed in Indonesia and Jamaica. Due to a delay in the procurement of tablets, CAPI could not be implemented in Rwanda. The results of the field tests are detailed in Chapter VI of this report.

# Methods of Crop Area Apportioning under Mixed Cropping and Intercropping

When more than one crop is sown simultaneously (i.e. crops are sown in mixtures), the fieldwise recording of area becomes challenging. To estimate the area under various crops, it is necessary to apportion the crop mixture area into the various component crops. The area can be apportioned by eye estimation or by means of a more objective method. Apportioning the area by eye estimation depends on the experience and judgement of the enumerator; this method may therefore introduce bias and lead to erroneous estimations. Apportioning area using objective methods such as measuring plant density, the row ratio (row intercropping), the width ratio (strip cropping) or the physical area occupied by each crop is therefore preferable, although expensive and time-consuming. For diagrams and examples of these cropping practices, see the Global Strategy's Working Paper titled *Synthesis of Literature and Framework – Research on Improving Methods for Estimating Crop Area, Yield and Production under Mixed, Repeated and Continuous Cropping*<sup>2</sup>. The objective methods are potentially capable of accurately reflecting the importance of each component crop in mixture, provided that the plant density of each of the component crops is sufficient. For example, in India, crops covering less than 10 per cent of the area in the mixture are ignored and the entire area is allocated to the main crop. This percentage threshold may vary from country to country.

---

<sup>2</sup> <http://gsars.org/en/synthesis-of-literature-and-framework-research-on-improving-methods-for-estimating-crop-area-yield-and-production-under-mixed-repeated-and-continuous-cropping/>.

**Table 2. Mixed cropping scenarios and respective methods of apportionment.**

<b>Crop mixture</b>	<b>Method of apportioning</b>
Temporary and temporary crops harvested at the same time Example: wheat and chickpea	Plant density
Temporary and temporary crop harvested in different seasons Example: sugarcane and green gram	Area double counted
Permanent and temporary crop Example: mango and sorghum	Area occupied by temporary crop recorded in harvested season and entire crop area possibly recorded under permanent crop
Permanent crop and permanent crop Example: mango and guava	Area may be apportioned on the basis of number of plants, with adjustment of the plant population of the pure crop of the component crop

In both objective and subjective measurement methods, the field area is apportioned between the component crops to “adjust” the observed area to the pure stand. In the following pages, each objective method is described in detail.

## **2.1. Area apportioning using an objective method**

### **Apportioning between temporary crops**

- **Using plant density**

When component crops are sown as mixture through broadcast sowing or narrow row spacing, the area under each component crop in the mixture may be apportioned on the basis of the adjusted plant density. The plant density per unit area for each component crop in the crop mixture is worked out on the basis of an objective method (i.e. counting the number of plants in a randomly selected plot), and the area of each component crop may be estimated by calculating the plant density ratio.

The plant density ratio is equal to the average number of plants of component crop *A* per unit area or the number of plants per unit area of component crop *A* in pure stand, divided by the average number of plants of component crop *B* per unit area or the number of plants per unit area of component crop *B* in pure crop stand.

For example, consider that there are three crops in the crop mixture over an area of 0.8 ha. Let there be 100 plants of Crop A when sown in crop mixture, and let there be 2,500 plants when Crop A is sown as pure. Let

there be 18 plants of Crop B when sown in crop mixture, and let there be 25 plants when Crop B is sown as pure. Let there be 80 plants of Crop C when sown in crop mixture, and let there be 200 plants when Crop C is sown as pure.

$$\text{Plant density ratio} = 100/2500 : 18/25 : 80/200 = 0.04 : 0.72 : 0.4$$

$$\text{Area of Crop A in the crop mixture} = (0.04/1.16) \times 0.8 = 0.0276 \text{ ha}$$

$$\text{Area of Crop B in the crop mixture} = (0.72/1.16) \times 0.8 = 0.4965 \text{ ha}$$

$$\text{Area of Crop C in the crop mixture} = (0.4/1.16) \times 0.8 = 0.2759 \text{ ha}$$

Apportioning the area under crop mixture on the basis of physical observation is expected to provide precise estimates of the crop areas of the mixture's component crops. It is suggested that several data points for each mixture be observed and then a fixed ratio worked out for each crop mixture, using the averaged value to apportion the mixture area to each component crop. The plant density of each component crop may be determined through physical observation, to be performed in a randomly selected experimental plot. The same exercise may be repeated for crops grown as pure. This exercise can be carried out when the crop-cutting experiments are being conducted.

It is suggested that the size and shape of the experimental plot for counting the plant population of the mixture's component crops could be a 5 m × 5 m square. The plot size for different food and non-food crops may be as follows:

Crop name	Shape	Length (m)	Breadth (m)	Diagonal (m)
Crop mixture, as a combination of two or more of any of the following crops: paddy, wheat, sorghum, pearl millet, finger millet, maize, groundnut, tobacco, sugarcane, green gram, chilli, horse gram, black gram, chickpea, sunflower and similar crops	Square	5	5	7.07

Once the experimental plot has been demarcated, the plant population of the experimental plot may be counted using the following method:

1. A piece of string should be tied tautly around pegs, which should be lowered gradually to the ground level.
2. The position of the string on the ground demarcates the boundary of the experimental plot.
3. The plants on this boundary are to be counted only if the roots fall by more than half within the experimental plot.

- **Using the row ratio** (*row intercropping*)

When the component crops are sown in separate rows, the area under each component crop may be apportioned on the basis of the row ratio of each component crop. The number of rows in a specified length is counted at three places chosen at random in the selected field, to determine the average number of rows.

- **Using the width ratio** (*strip cropping*)

In intercropping, where each component crop is sown in separate distinct groups of rows called “strips”, the width of the strip of each component crop is measured at three points, selected at random. The width ratio between the component crops is determined on the basis of the average width of the component crop in the strip intercropping. Therefore, the area of component crops can be apportioned using the width ratio of the component crops.

### **Apportioning between permanent and temporary crops**

The apportioning of crop area between permanent crops (for example, fruit trees) and temporary (seasonal and annual) crops under mixed cropping is a complex and controversial operation. Generally, temporary crops are grown in the orchards until the bearing stage; after bearing, they are grown in the spaces between the permanent crops. Shade-loving crops are also grown in the orchards. The permanent crops are planted once and harvested every fruiting season. The permanent crops remain in the field for a longer period of time, occupying the entire area. The temporary crops occupy a certain area of the orchard; therefore, the actual area occupied by the temporary crops may be allocated to the temporary crop area for estimating the temporary crops’ production, except for cases in which the area under temporary crops is lower than the threshold defined by National Statistical Offices (NSOs). In these cases, the temporary crops can be ignored.

If more than one temporary crop is grown in the mixture, then the mixture area should be apportioned between the temporary crops.

The production estimates of permanent crops (mango, guava, apple, etc.) are based on the per tree yield and the number of trees.

The orchard's entire area may be allotted to the permanent crop. The area under the temporary crop sown in orchard may be apportioned on the basis of the average canopy area ( $\pi r^2$ ) of three to five randomly selected trees in the orchard, where  $r$  is the radius of the canopy. Multiplying the average canopy area per tree by the total number of trees gives the total area of the orchard's permanent crop. This area may be deducted from the orchard's total area and the remaining area may be allocated to the temporary crop grown in the orchard. If more than one temporary crop is grown in the orchard, the remaining area under the temporary crops can be apportioned using the physical observation method. However, the method of apportioning may underestimate the crop area, thereby leading to an overestimated yield. For example, if mango or coffee is the permanent crop and the recommended spacing between trees (even in pure stand) is greater than the sum of the canopy areas, then this method would underestimate the area occupied by the permanent crop and thus overestimate the yield.

With regard to shade-loving crops that are grown under the canopy, the area may be counted under the temporary crop during their respective harvesting seasons.

### **Between permanent crops**

In case of mixtures of two or more permanent crops, the area may be apportioned on the basis of the number of trees of each component crop by adjusting the area to the pure crop. The number of trees or plants of each permanent crop may be recorded separately.

## **2.2. Area apportioning using a subjective method**

- *Apportioning using seed rates.* When component crops are sown as a mixture without any row arrangement, the area under each component crop in the mixture may be apportioned on the basis of the adjusted seed rate. In these cases, only the seed used is considered when apportioning the area; the population of plants in the field is not taken into account. The seed sown may not have fully germinated, or the plants may show a less-than-optimal survival rate. However, the area apportioned to each

crop on the basis of seed rate may lead to biased estimates if only some of the seeds sown have germinated.

If  $a$  kg and  $b$  kg are the quantities of seeds used for sowing a mixture of two component crops, and  $A$  kg and  $B$  kg are their normal seed rates when sown as pure crops, the proportion of area under each component crop may be estimated as  $a/A : b/B$ . The normal seed rates are determined by using seed rate information in pure stand from five randomly selected field in a subdistrict and taking the average value of these.

For example, consider that the mixture contains two crops,  $A$  and  $B$ .

Consider further that the area under crop mixture is 0.4 ha, the quantity of seed used to sow crop A in crop mixture A and B is 50 kg, and that the normal seed rate of crop A is 120 kg/ha. The quantity of seed used to sow crop B in crop mixture A and B = 1 kg, and the normal seed rate of crop B is 5 kg/ha. Then, the seed rates ratio is equal to  $(50/120) : (1/5) = 0.42 : 0.2$ .

The area of crop A in crop mixture A and B is  $(0.42/(0.42+0.2)) \times 0.4 = 0.27$  ha

The area of crop B in crop mixture A and B =  $(0.2/(0.42+0.2)) \times 0.4 = 0.13$  ha. In mixed cropping, when the crop area is apportioned on the basis of seed rate, the farmer may incorrectly report the seed rate. Further, the size of the seed (or the test weight) may influence the area of each component crop, which may result in the incorrect apportioning of the crop area.

The seed rate method was implemented in the field tests.

- *Ignoring intercropping.* In this method, crop areas are recorded only for crops grown in pure stand. This leads to the underrepresentation of the actual area. The crop yields are therefore overestimated (Fermont & Benson, 2011).
- *Recording only main crop:* In this method, crop areas are recorded only for the major crop in the mixture. The crop area and yield estimates are reported as if they were obtained from crops grown in pure stand. Thus, the total area of a particular crop is estimated as the sum of the total area of the crop grown in pure stand and the total area in which the crop is grown as a major crop. The crop yield is determined where crop is grown as a pure stand or as a major crop (Fermont & Benson, 2011).

- *Using the whole plot as a denominator for each crop in the mixture:* According to this method, the crop area under mixture is double-counted, even within the same season. The total area for a crop consists of the area under the crop in pure stand and the entire plot area in which the crop is grown in mixture. The average yield for crops is determined separately in pure stand and in mixture. This method will lead to the overestimation of crop areas; however, it is used in some European countries (Fermont & Benson, 2011).
- *Using a fixed area ratio.* The apportionment of the area under each crop may be performed at a higher level (e.g. district or subdistrict level) using a fixed-area ratio determined through subjective eye estimates carried out at a periodical interval. This method of apportionment of the crop area into component crops is followed in some states of India, for the recognized mixtures (India, 2008).
- *Ignoring crops occupying less than the threshold level (as determined by the relevant NSOs) of the plot area in crop mixture.* Crops that are grown in the mixture in an extremely low proportion, for example less than the threshold level, may be ignored. Thus, in cases where there are two crops in the mixture, the entire area may be considered as a pure crop.
- *Dividing total field area sown equally between each component crop.* This is a simple method; however, it will lead to the over- and underestimation of crop production.
- *Allocating total area sown to each component crop in the mixture.* This method is crude and may therefore lead to overestimating the areas.
- *Allocating area when component crops are harvested in different seasons.* When temporary crops (seasonal and annual crops) are sown in crop mixture at the same time and harvested during different seasons, the entire area of mixture is treated as double-cropped. The whole area is recorded under each component crop in the respective seasons during which the crop is harvested. For example, in some countries, corn is harvested after seven months, while beans are harvested after three months. This implies that these two crops are harvested during two different seasons and, therefore, that the area under corn and bean is double-counted.

Simple additions to existing Crop-Cutting Experiment (CCE) questionnaires will enable countries to collect the information required to determine the area under mixed crops using physical observations of plant density, row ratio, seed rates, etc. (refer to the Field Test Protocol, Sections 4.0 to 6.0 of Questionnaire CCE I, for the inclusion of relevant questions).

# Area Measurement Methods and Data Collection

## 3.1. Introduction

This chapter describes various subjective and objective methods used to determine crop area and their respective advantages and disadvantages. Chapter VI of this report summarizes the results of experiments carried out by the Global Strategy and the World Bank to compare these methods.

## 3.2. Measurement methods used for crop area estimation

Crop area plays an important role in estimating crop production. The accuracy of crop production estimates depends on the accuracy of crop area estimates. The most appropriate measurement technique to estimate crop area depends on various operational factors, such as land configuration, field shape, crop type, cropping pattern, available skills and resources (Casley & Kumar 1988). Crop area may be estimated either directly, by means of measurements, or by visual estimation. This section describes the various methods used to determine crop area, as well as their respective advantages and disadvantages. Both subjective and objective methods are considered.

### 3.2.1. Farmer assessment of crop area

In this method, the farmers are asked to estimate the area of their fields. The enumerator and the farmer may visit all of the farmer's fields and estimate the surface area by visual inspection (David, 1978). Notably, if some plots are located far apart from each other, the farmer can declare the size of the area without necessarily having to visit the plot with the enumerator. The results of the field tests conducted in Indonesia and Rwanda show that the method can provide satisfactory estimates of parcel area for small parcel sizes. However, the results from Jamaica are not as encouraging. The evidence shows that the farmer assessment method is workable in countries where farmers are aware of the units of area measurement (Verma *et al.*, 1988). The method is therefore likely to provide useful results where the mixtures of crops are at the same stage of growth or where systematic intercropping is used.

## **Advantages**

This method is relatively less time-consuming and inexpensive. Farmer assessment does not require the enumerator to visit the individual plots, which is cost-effective particularly if the plots are located far away from the location of the initial interview. Furthermore, farmer assessments of crop area can serve as a baseline for imputation where objective measurements are missing (Kilic et al., 2013).

## **Disadvantages**

This method is highly subjective, as it depends on farmers' knowledge and experience. Furthermore, any nonstandard units of measurement used by farmers may be difficult to standardize. The farmers may also have incentives to misreport crop area for reasons such as taxation. The data analysis conducted within the World Bank's study of four African countries (Carletto et al., 2015) indicate that self-reported land areas systematically differ from GPS land measurements, and that this difference leads to biased estimates of the relationship between land and productivity and consistently low estimates of land inequality. Furthermore, results from methodological experiments carried out by both the World Bank and the Global Strategy indicate that farmers tend to overreport plot area for small plots, and underreport area for very large plots.

### **3.2.2. GPS**

GPS is a space-based satellite navigation system that provides location and time information anywhere on Earth. GPS hardware determines coordinates for the  $x$ ,  $y$  and  $z$  axes, with  $x$  and  $y$  being the geographic coordinates that determine location and  $z$  being the coordinate that determines elevation. Initially, GPS was used to determine the location of a particular point. However, with advancements in technology, it is now capable of determining the elevation and even the area covered. As a result, GPS has become a very important tool for measuring the area under a crop, with the added advantage of requiring reduced time and labour.

## **Advantages**

Area measurements with GPS are more rapid, time-efficient and feasible. In addition, they are in digital format, and thus traceable and easy to incorporate into a database. One major advantage of GPS, as with any objective measurement, is that it is immune to the potential biases linked to respondent characteristics and the use of non-standard measurement units (Carletto *et al.* 2016b). In three field-testing countries, the area measured by GPS was used as the gold standard for comparing other measurement methods. The World Bank study reports that the more systematic use of GPS-measured land area may result in improved agricultural statistics and a more accurate analysis of agricultural relationships (Carletto *et al.*, 2016a).

## **Disadvantages**

The accuracy of GPS measurements is influenced by (i) the tree canopy cover (accuracy is high with no tree canopy cover and lower with partial or dense tree canopy cover); (ii) the weather conditions (accuracy is higher under sunny conditions than under cloudy conditions); (iii) the plot size (the larger the size of the plots, the more accurate the results); and (iv) the land in hilly areas. Securing ample power supply is one of the major problems faced when using a GPS device for measurement, as is travelling to the plot to take the measurement required. As a result, data is commonly missing when plots are located in remote areas that are difficult for the enumerator to reach (Carletto *et al.*, 2016a).

### **3.2.3. Area measurement through maps**

This method involves the preparation of orthophotography and/or high-resolution satellite imagery, and the enumerator drawing the plot boundaries directly on the map. Sometimes, the plot boundary is visible and can be easily drawn on the map. However, in most cases, enumerators use measuring tape to measure the size of the plot and, using the map scale, then draw the plot on the map. To draw plots accurately, triangulation can be used. Screening is required before the maps are digitalized. The plot area can be calculated from digitalized maps with any Geographic Information System (GIS) software.

## Advantages

This method can provide complete coverage and accurate measurements if the satellite image is of high quality and up-to-date.

## Disadvantages

The acquisition of orthophotographs and digitized maps can be expensive, although the costs are gradually declining. To accurately determine plot area, maps must be updated on the basis of remote sensing satellite imagery, because plot boundaries may change due to the combination of two or more plots into a single plot, or a division split of one plot into two or more plots. This method also requires clear satellite imagery, which may not be possible to obtain due to weather conditions.

### 3.2.4. Rope-and-compass method

This method, also known as the polygon method, traverse measurement, traversing, chain-and-compass, or Topofil method, is one of the most prevalent traditional methods used to measure crop area (Schøning *et al.*, 2005). Until GPS methods became available, it was considered the **gold standard** for crop area estimation, in view of its potential to provide accurate area figures. Where the plots are of a regular shape, the method involves measuring the length of each side and the angle of each corner using a measuring tape and a compass. The plot's surface area can then be calculated using trigonometry (FAO, 1982). For irregularly shaped plots, an approximate polygon with straight sides is obtained by demarcating its vertices on the ground. Due care is taken to balance the protruding pieces left out from the process by including other small pieces that are not part of the plot. During the give-and-take process and the measurement process, errors are introduced. According to Casley & Kumar (1988), if the polygon does not close and the closing error exceeds 3 per cent of the polygon's perimeter, the measurement procedure should be repeated.

In this method, the boundaries of a field to be measured are first identified by use of sight poles, and taking compass bearings and measuring the length of each side of the polygon obtained. FAO's Statistics Division has developed several methods for calculating areas with programmable calculators (FAO, 1982).

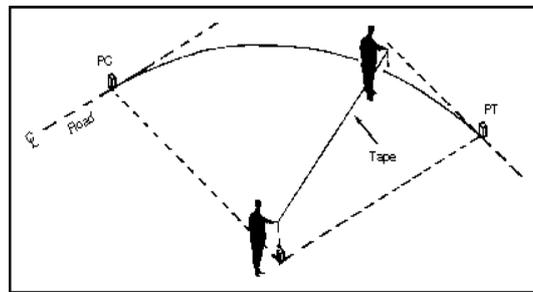
## Advantages

This method often provides accurate area measurements and can be used directly in the field when measurements are made (FAO, 1982). The closure error can be evaluated on the spot, and when the error of the measurement is considered to be too great, the process can be repeated.

## Disadvantages

Obtaining area measurements through this method is laborious, time-consuming, and expensive. At least two enumerators are required for each plot.

Figure 3.2.4.1. Area measurement using rope and compass.



## 3.2.5. Remote sensing and GIS

Remote sensing and GIS technology have been widely adopted to estimate crop area statistics. For this purpose, classified satellite images and land cover maps produced by photo-interpretation are useful tools. It is not recommended to directly use satellite images (in terms of pixel counting) for the area measurement or simple area measurement of polygons of a land cover map.

Initially, two broad approaches to the use of remote sensing to generate crop statistics were recognized:

1. Direct and independent estimation that uses remote sensing data and a recognition technique to estimate the crop area in the study region.

Use of remote sensing data as an auxiliary variable, to help enhance the precision of the estimates based on ground surveys and reduce the amount of field data to be collected, if the precision to be reached has been fixed; if the sample size is fixed, this approach provides more precise estimates.

### **Advantages**

This method provides quick crop area estimates covering a vast geographical area. It is also useful for obtaining estimates of areas in hilly terrains and in areas that are inaccessible.

### **Disadvantages**

The method is expensive. There may be problems in obtaining estimates for areas under cloud cover. The area estimates may not be accurate for small plots. However, the method may be satisfactorily used to determine plot area in countries where plots tend to be very large (e.g. the United States).

# Yield Measurement Methods and Data Collection

## 4.1. Introduction

This chapter describes various methods used to determine crop yield and their respective advantages and disadvantages. Both subjective and objective methods are considered. It is recalled that Chapter VI of this Technical Report summarizes the results of the experiments conducted by the Global Strategy and the World Bank to compare these methods.

## 4.2 Measurement methods for yield estimation

Estimation of crop yield is always a challenging exercise, which is further compounded when crops are mixed, yield estimation is carried out in farms owned by smallholders, or there is no cadastral information on land use (Murphy, Casley & Curry, 1991). Although subjective methods to determine crop yield lack the capacity to produce accurate estimates of yield rates, countries prefer to use these methods because they are easy to implement. On the other hand, the objective methods to determine crop yield are costly and difficult to implement; however, they are capable of providing accurate estimates. This section describes various subjective and objective methods that are commonly used to determine crop yield, along with their respective advantages and disadvantages.

### 4.2.1. Whole plot harvest

The whole plot harvest method is employed in detailed farm surveys and in demonstration plots (Norman et al., 1995). This method is regarded as the absolute standard for crop yield estimation, especially if applied together with the farmer (Casley & Kumar, 1988).

#### **Advantages**

The main advantage is that it is almost bias-free, as all sources of upward bias reported for crop cuts can be eliminated when the whole field is harvested. This

method is suitable for small-scale investigations of a case-study nature (Poate, 1988). Complete harvesting generates more accurate data than crop cuts, because the bias from within-field variability – which is commonly 40 to 60 per cent of total yield variability – is eliminated.

### **Disadvantages**

The main drawback of the method is that it involves a large volume of work, making it unsuitable for moderate and large sample sizes or multiple crop studies.

### **4.2.2. Crop cut method**

The crop cut method was developed in the late 1940s in India for estimating crop yield on the basis of the sampling of small subplots within cultivated fields. It was created by pioneers in the field of sampling and survey design: P.C. Mahalanobis of the Indian Statistical Institute and P.V. Sukhatme of the Indian Council of Agricultural Research (ICAR). The method involves the random demarcation of a plot of a specified size and shape, harvesting the produce from the plot, and threshing, winnowing and drying the produce to determine its dry weight.

### **Advantages**

Since being endorsed by FAO in the 1950s, the crop cut method has been commonly regarded as the most reliable and objective method for estimating crop yield. A sufficient number of cuts in a sufficient number of fields provides a valid estimate of average yield (Murphy et al., 1991). Another advantage of the crop cut method is that the productivity of parcels, subparcels or fields can be determined without knowledge of their size.

### **Disadvantages**

The crop cut method measures the biological yield, which does not necessarily take into account harvest losses and therefore does not reflect the economic yield that is of use to the farmer or planner. However, certain countries, such as the United States, take into account harvest loss at the time of crop cutting<sup>#</sup>. Obtaining yield estimation through crop cuts is both time-consuming and labour-intensive. To facilitate fieldwork and reduce costs and time required, a clustered sampling procedure is usually applied when crop cuts are used for larger-scale

surveys. The results of all of the field tests show that this method tends to overestimate field production.

#### **4.2.3. Farmer recall**

This method of post-harvest estimation is commonly performed at the farmer's house or at the site where the harvest is stored, for the enumerator to cross-check the estimates with the available storage capacity (Casley & Kumar, 1988). Depending on rainfall distribution, the recall period may range from six months or one season to three years, or three to six seasons (Howard et al., 1995; Lekasi et al., 2001; Erenstein et al., 2007). The method has the potential to provide accurate estimates of crop production in countries that have achieved higher levels of mechanization, commercialization and record-keeping (Hagblad, 1998). It is useful where farmers are literate and knowledgeable (Kelly et al., 1995; Casley and Kumar, 1988).

##### **Advantages**

The method is simple, the data are quickly available, and is less expensive to implement. The method can be used as an auxiliary variable in crop yield estimation.

##### **Disadvantages**

The method is subjective and likely to yield inaccurate data if the recall period is very long (Howard et al., 1995). It is useful for determining crop production. Therefore, the availability of accurate estimates of crop area is a prerequisite for determining crop yield. Some of the method's shortcomings are (i) ignorance of in-kind payments; (ii) non-standard harvest units; (iii) intentional over- or underreporting; (iv) low accuracy with longer recall periods; (v) historical average production factors; (vi) poor quality responses in lengthy interviews; (vii) insufficient supervision; and (viii) illiteracy, especially in African countries, which results in inaccurate responses (David, 1978; Casley & Kumar, 1988; Poate, 1988; Rozelle, 1991; Howard et al., 1995; Kelly et al., 1995; Diskin, 1997; UBOS, 2002; Ali et al., 2009; Fermont & Benson, 2011).

#### **4.2.4. Farmer prediction**

This method of pre-harvest estimation is commonly performed on a plot-by-plot basis, and both the enumerator and the farmer are in visual contact with the growing crop. The method is useful when it is used to predict crop production

15 days before harvest. The results of the field test conducted in Indonesia reveal that the farmer prediction method exhibits a high correlation with the CCE and sampling of harvest unit methods. The method is useful where farmers are literate and knowledgeable (Kelly et al., 1995; Casley & Kumar, 1988).

### **Advantages**

The use of farmer prediction is not particularly laborious (Murphy et al., 1991; Casley & Kumar 1988). In comparison to the crop cut method, farmer estimation is less costly and faster to carry out. Consequently, farmers' estimations with the same resources allow for a larger number of yield estimates to be collected, than do crop cuts. This method is a valuable source of auxiliary information if problems are encountered in crop yield estimation.

### **Disadvantages**

Some of the method's shortcomings are: (i) use of non-standard harvest units; (ii) intentional over- or underreporting; (iii) use of historical average production factors; (iv) poor quality responses in lengthy interviews; (v) insufficient supervision; and (vi) illiteracy, especially in African countries, which results in inaccurate responses (David, 1978; Casley & Kumar, 1988; Poate, 1988; Rozelle, 1991; Howard et al., 1995; Kelly *et al.*, 1995; Diskin, 1997; UBOS, 2002; Ali *et al.*, 2009; Fermont & Benson, 2011). Several studies indicate that the use of farmers' estimates is affected by the bias in estimation. The use of this method as a source of auxiliary variables for crop yield estimation lacks consistency, as evident from the field test results. Furthermore, farmers are only capable of predicting the crop produce in local units; this requires local units to be converted into standard units.

---

# The procedure devised by the USDA involves carrying out a final crop cutting at maturity or immediately before harvest. "Sample fruit (ears, pods, bolls, heads, or tubers) is sent to a lab to determine fruit weight, threshed grain weight, and moisture content. A postharvest visit is made to glean fruit left in some sample fields." USDA-NASS. "Objective Yield". [https://www.nass.usda.gov/Surveys/Guide\\_to\\_NASS\\_Surveys/Objective\\_Yield/index.php](https://www.nass.usda.gov/Surveys/Guide_to_NASS_Surveys/Objective_Yield/index.php). Accessed 26 January 2016.

#### 4.2.5. Sampling of harvest units

In this method, instead of harvesting and weighing the whole field, the enumerator may wait for the farmer to harvest his or her field. An attempt is made to estimate the number of units – e.g. sacks, baskets and bundles – harvested by the farmer. The enumerator then randomly selects a number of harvest units and weighs them to obtain an average unit weight. The harvest units are generally sampled immediately prior to storage, and include a measurement of the harvested product's moisture content (Casley & Kumar, 1988). The results of the field test in Indonesia on using this method in crop yield estimation are encouraging. However, the utility of the method is limited, as evident from the field test carried out in Jamaica. The good performance in Indonesia may be due to the fact that the parcel size was generally smaller than in Jamaica. Further, only root crops were studied in Jamaica; for these, the harvest unit method of sampling is considered unsuitable (Fermont & Benson, 2011).

##### **Advantages**

The technique is straightforward and can be used on larger samples, compared to the crop-cut and whole-plot harvesting methods. Unlike in the case of farmer estimates, it does not matter if the harvest units are particular to each individual farmer, because the enumerator either weighs the complete harvest or weighs a random unbiased selection of the harvest units of each farmer (Poate, 1988).

##### **Disadvantages**

When the harvest is stored in one or several large granaries or stores, the enumerator must apply analytical skills to accurately estimate total production (Rozelle, 1991). This method is considered unsuitable for crops with an extended harvest period and multiple pickings, such as root crops, banana, cotton and similar crops. The method lacks consistency in providing accurate yield estimates.

#### 4.2.6. Expert assessment

Experts that have extensive experience with crops, such as extension staff, field technicians or subject matter experts, can estimate crop yield by either visually assessing the field or by estimating yield, combining tools such as visual assessment, field measurements, and empirical formulas (Fermont & Benson, 2011). This technique provides an estimate of biological yield. The method could

not be tested in the field-testing countries. However, the results of the field test in Rwanda revealed that the method of enumerator assessment of crop produce has the potential to provide satisfactory estimates of crop yield. Similarly, the results in Jamaica revealed a promising performance of the method of farmer assessment (by eye estimate) of crop produce on the day of harvest.

### **Advantages**

An advantage of the expert assessment method are that it can be applied on a relatively large scale, compared to the crop-cut and farmer estimation methods; in addition, it does not require area estimation and eliminates a source of potential bias. Other important advantages are that one team of experts can be used throughout a study, which results in a similar bias for all yield estimation (Rozelle, 1991), and it is cheaper to implement than other methods.

### **Disadvantages**

Eye estimations of crop yield require not only practical but also technical familiarity with the yield potential of different varieties of a crop and their relative performance in different environments (David, 1978). The accuracy of the yield assessment, therefore, strongly depends on the expert's level of expertise. When assessments are made by extension officers, yield estimation may be biased upward, especially if the assessments are made in their own work area and the information collected thus pertains to the quality of their own work (Casley & Kumar, 1988). In contrast, Bradbury (1996b) reported that yield estimates by means of expert judgment in Europe were generally considered to be biased downward. Considering that a national survey or an agricultural census requires yield estimates of a large range of crops, it is difficult to identify experts that possess the practical and technical expertise required to provide accurate estimations across all crops.

#### **4.2.7. Crop diary and crop diary with telephone calls**

In this method, diaries are given to farmers for recording the crop produce on a continuous basis. In the method of crop diary with telephone calls, in addition to the crop diary, the enumerator makes two telephone calls per week to ensure that the the farmers properly make the recordings in the diary. The crop diary method is useful to capture produce from crops with extended harvest periods, such as cassava, banana and sweet potato., because farmers may encounter problems in remembering the amounts harvested over time for one or several plots. The method was extensively used in the Living Standard Measurement Study

(LSMS) in Zanzibar, Tanzania under the Measuring Cassava Productivity (MCP) study (Carletto *et al.*, 2016a).

### **Advantages**

The method is cost-effective and provides reliable yield estimates of crops with an extended period of harvest.

### **Disadvantages**

Illiterate farmers may find it difficult to fill the diary.

### **4.2.8. Crop cards**

The crop card method is a refined version of the farmer recall method. It also estimates the economic yield. The method was evolved to obtain more reliable yield estimates of crops with an extended period of harvest, e.g. cassava, banana and sweet potato, because farmers may have difficulties remembering the amounts they harvested over time for one or several plots. Under this method, each farmer in the survey is given a set of crop cards by a Crop Card Monitor (CCM) and receives training on how to use them to record the quantity that the farmer harvested in local harvesting units after each harvest operation. The CCM is expected to visit each farmer on a regular basis, to monitor the farmers' recordings and to correct any problems the farmer may have. Then, after a certain period, the CCM collects all cards for processing.

This method was tested in Uganda during the Uganda National Household Survey of 2005-2006 and was compared with farmer recall estimates. Further, using the data collected for UNHS 2005-2006, Carletto *et al.* (2010) showed that crop card production estimates were 40 to 60 per cent lower than the farmer recall production estimates for both crops with an extended harvest time (cassava and banana) and for other crops (maize and beans). This was in line with the findings of Sempungu (2010), who, using the same data set, found that cassava and sweet potato yield estimates from the crop card method were, respectively, 30 and 46 per cent lower than those obtained from farmer recall. The above studies suggested, first, that farmers were either seriously overestimating crop production during the recall exercise or underestimating crop production with the crop card method and, second, that the upward or downward bias that resulted does not seem to depend on the type of crop. This contradicts the assumption that farmers have difficulties in accurately recalling multiple harvests of crops over an extended harvest period.

## **Advantages**

This method provides more reliable yield estimates of crops with an extended period of harvest than the farmer recall method, as farmers find it difficult to remember the amounts they harvested over time for one or several plots.

## **Disadvantages**

This method presents several problems, including irregular monitoring by enumerators, illiterate farmers who are incapable of filling in the crop cards, some recordings that included crop purchases, and a very large range of observed harvest units (Ssekiboobo, 2007).

### **4.2.9. Crop modelling**

This method is widely used to estimate average biological yield in the case of smallholder farmers. Crop models vary widely in their complexity. The simplest sets of models are of empirical-statistical nature, whereas the most complex models are based on crop physiology. The former aims to find the best correlation between crop yield and environmental factors such as weather parameters (temperature, humidity, rainfall, etc.) from long-term data sets. Using the established relations, the model attempts to predict crop yield at a regional or national level on the basis of actual environmental observations, whereas crop growth models estimate crop yield as a function of physiological processes and environmental conditions. They range from relatively simple models that take into account only basic crop physiology processes (e.g. Penman-Monteith models based on the estimation of actual evapotranspiration) to extremely complex models that estimate daily gains in biomass production by taking into account all known interactions between the environment and physiological processes (Sawasawa, 2003). The crop modelling approach is used in India for multiple season crop forecasting, utilizing weather parameters as well as parameters such as crop area and price in previous years, under the project entitled *Forecasting Agricultural output using Space, Agro-meteorology and Land based observations* (Parihar & Oza, 2006).

### **Advantages**

Crop models can be used to predict crop yield in specific conditions or a range of conditions, and are an extremely useful tool in research studies exploring the impact of specific factors on average crop yield.

### **Disadvantages**

Crop models cannot be used to predict crop yield for individual farmer fields, as this requires a far too great amount of input data.

# Methodology for Estimating Crop Area and Crop Yield under Mixed and Continuous Cropping

## 5.1. Introduction

The purpose of this chapter is to describe the methodology proposed for estimating crop area and crop yield in the context of mixed and continuous cropping. Both list and area frames are considered. The sample selected for area estimation is to be used as the sampling frame for selecting the sample for crop yield estimation. The sampling design used to select the sample for crop area and yield estimation is explored briefly. The subjective and objective methods are combined using a double sampling regression estimator. Estimation procedures based on the domain estimation approach, using a double sampling regression estimator, for crop area and yield estimation are also seen. The theory of domain estimation allows for the separate estimation, from a single sample, of crop area and yield of different crops and of their mixtures. In addition, the criterion for determining sample size is also provided.

## 5.2. Sampling design used for crop area estimation in mixed cropping

In this section, the sampling design used for estimation of crop area under mixed cropping is described for two different scenarios: (1) adoption of the area frame approach and (2) adoption of the household approach.

### 5.2.1. Description of the area frame approach

Area frames may consist of an infinite set of points or of a finite set of area segments. The segments of an area composing an area frame can be determined in different ways: they may be established by reference to identifiable physical boundaries, such as rivers or roads, by means of a square grid of map

coordinates, or by making their limits coincide with those of agricultural holding lands (FAO, 1996). When the segment does not coincide with the boundaries of a holding, a tract must be defined. The segments are then subdivided into non-overlapping tracts, in which a tract is the part of a holding that is found within the limits of a segment, or a piece of land that does not belong to any holding. A holding comprises of at least one tract. Tracts are observational units. GSARS (2015b), Gallego *et al.* (1994) and Gallego (1995, 2013) provide information on sampling points from an area frame for agricultural surveys.

One of the main advantages of area frames is that they provide full coverage of the target population and do not present duplication. Further, once an area frame is constructed it remains up-to-date for a long time. Area frames can be applied to generate estimates of parameters of land areas, such as the total cultivated area, as they enable the recording of objective measurements on the ground. The presence of outliers in samples from area frames has a considerable impact on estimates (Carfagna, 2004). For the purposes of selecting the sample, a stratified two-stage cluster sampling design with two phases during each stage was employed, using the available area frame. In the following sections, we provide details on the estimation procedure.

### 5.2.1.1. Estimation procedure for crop area estimation under mixed cropping when the area frame is available

Let

$H$  = number of subdistricts in a district that can be considered as  $H$  strata  
 $N_h$  = number of Enumeration Areas (EAs) in the  $h^{\text{th}}$  subdistrict (stratum) that are considered PSUs  
 $h = 1, \dots, H$ .

In general, the total number of EAs in each subdistrict,  $N_h$ , is known. Let

$M_{hi}$  = number of segments (SSUs) in  $i^{\text{th}}$  EA of  $h^{\text{th}}$  subdistrict  
 $T_{hij}$  = number of parcels ultimate sampling unit in the  $j^{\text{th}}$  segment of  $i^{\text{th}}$  EA in the  $h^{\text{th}}$  subdistrict.

In these parcels, crops are grown in different types of mixture (e.g. pure stand, mixture-1 or mixture-2). In this case, therefore, the different crop mixtures are taken as the domains of the study. It is assumed that in each  $h$  subdistricts,  $D$  different crop mixtures are followed as pure stand, mixture-1, mixture-2 etc. Thus, there would be  $\{U_{h1}, \dots, U_{hd}, \dots, U_{hD}\}$  domains in the  $h^{\text{th}}$  stratum.

Let

$y_{hijk}$  = crop area of  $k^{\text{th}}$  parcel (USU) within  $j^{\text{th}}$  segment (SSU) of  $i^{\text{th}}$  EA (PSU) in  $h^{\text{th}}$  subdistrict (stratum).

The total area under the  $d^{\text{th}}$  crop mixture (domain) in a district is given by

$$Y_d = \sum_{h=1}^H \sum_{i=1}^{N_{hd}} \sum_{j=1}^{M_{hid}} \sum_{k=1}^{T_{hijd}} y_{hijk} .$$

The population total based on all domains is given as

$$Y = \sum_{d=1}^D Y_d = \sum_{d=1}^D \sum_{h=1}^H \sum_{i=1}^{N_{hd}} \sum_{j=1}^{M_{hid}} \sum_{k=1}^{T_{hijd}} y_{hijk} .$$

In this situation, the proposed sampling design for estimating crop area at district level is the stratified two-stage cluster sampling design; there are two phases to each stage of sampling and the area frame approach is used.

Let, in the first phase,

$n_h'$  = number of EAs selected from  $N_h$  EAs (PSUs) by probability proportional to size with replacement design from the  $h^{\text{th}}$  stratum. The probability of selecting  $i^{\text{th}}$  EA in  $h^{\text{th}}$  stratum is computed with  $z'_{hi} = X_{hi}/X_h$ , where  $X$  indicates the total agricultural land.

In the second phase of the first stage, let

$m_{hi}'$  = number of segments selected from  $M_{hi}$  segments, by Simple Random Sampling without Replacement (SRSWOR) design in the  $i^{\text{th}}$  EA of  $h^{\text{th}}$  strata.

$T_{hij}$  = number of parcels and growing crop mixtures that are completely enumerated for collecting auxiliary information on the parcel such as seed used, farmer assessment in the  $j^{\text{th}}$  segment of  $i^{\text{th}}$  EA of  $h^{\text{th}}$  strata

$n_{hd}'$  = number of EAs of selected  $n_h'$  EAs growing specific  $d^{\text{th}}$  mixture

$m_{hid}'$  = number of selected segments of  $m_{hi}'$  selected segments growing  $d^{\text{th}}$  mixture

$T_{hijd}$  = number of parcels of  $T_{hij}$  parcels growing  $d^{\text{th}}$  mixture

In the second stage of sampling, let

$n_h$  = number of EAs selected from  $n_h'$  initially selected EAs (PSUs) by SRSWOR

$m_{hi}$  = number of segments selected by SRSWOR from  $m_{hi}'$  segment and in 40 per cent of randomly selected tracts within a selected segment; all  $T_{hij}$  parcels in the selected tracts are to be completely enumerated. The area of each parcel growing crop mixtures of these sampled segments is measured by GPS. The data collection is to be performed using the questionnaires provided in Annex B of the Field Test Protocol document. The area for component crops is obtained by apportioning, using the information on seed rates.

$n_{hd}$  = number of EAs of  $n_h$  randomly selected EAs growing specific  $d^{\text{th}}$  mixture

$m_{hid}$  = number of segments growing  $d^{\text{th}}$  mixture in  $m_{hi}$  sampled segments

$T_{hijd}$  = Number of parcels in which  $d^{\text{th}}$  mixture is grown out of  $T_{hij}$  parcels

The aim is to estimate the total crop area under a specific crop ( $Y$ ) and under different mixtures ( $Y_d$ ),  $d=1,2,\dots,D$ .

Let

$x_{hijk}$  = auxiliary information (e.g. seed used or farmer assessment) corresponding to  $k^{\text{th}}$  parcel (SSU) of  $j^{\text{th}}$  selected segment in  $i^{\text{th}}$  EA (PSU) within  $h^{\text{th}}$  subdistrict (stratum)

$y_{hijk}$  = crop area of corresponding parcel, measured by GPS

The double sampling regression estimator of the total area under the  $d^{\text{th}}$  mixture under the stratified two-phase two-stage cluster sampling design can be computed with

$$\hat{Y}_{lr2d} = \hat{Y}_d + b_{A2d}(\hat{X}_d' - \hat{X}_d) \quad (2)$$

where

$$\begin{aligned} \hat{Y}_d &= \sum_{h=1}^H \frac{1}{n_h} \sum_{i=1}^{n_{hd}} \frac{1}{z_{hi}'} \frac{M_{hi}}{m_{hi}} \sum_{j=1}^{m_{hid}} \sum_{k=1}^{T_{hjd}} y_{hijk} = \sum_{h=1}^H \frac{1}{n_h} \sum_{i=1}^{n_{hd}} \frac{1}{z_{hi}'} \frac{M_{hi}}{m_{hi}} \sum_{j=1}^{m_{hid}} y_{hij.d} = \sum_{h=1}^H \frac{1}{n_h} \sum_{i=1}^{n_{hd}} \frac{\hat{Y}_{hid}}{z_{hi}'}, \\ \hat{X}_d &= \sum_{h=1}^H \frac{1}{n_h} \sum_{i=1}^{n_{hd}} \frac{\hat{X}_{hid}}{z_{hi}'}, \quad \hat{X}_d' = \sum_{h=1}^H \frac{1}{n_h} \sum_{i=1}^{n_{hd}} \frac{\hat{X}_{hid}'}{z_{hi}'}, \\ \hat{Y}_{hid} &= \frac{M_{hi}}{m_{hi}} \sum_{j=1}^{m_{hid}} \sum_{k=1}^{T_{hjd}} y_{hijk}, \quad \hat{X}_{hid} = \frac{M_{hi}}{m_{hi}} \sum_{j=1}^{m_{hid}} \sum_{k=1}^{T_{hjd}} x_{hijk}, \quad \hat{X}_{hid}' = \frac{M_{hi}}{m_{hi}} \sum_{j=1}^{m_{hid}} \sum_{k=1}^{T_{hjd}} x_{hijk}' . \end{aligned}$$

It should be noted that the double sampling regression estimator is useful when there is a high correlation between the subjective and objective methods of determining crop area. Establishing the extent of correlation between the subjective and objective methods is a prerequisite to developing the double sampling regression estimator.

By minimizing the variance of the linear regression estimator  $\hat{Y}_{lr2d}$  with respect to  $b_{A2d}$  and ignoring the variation in  $b_{A2d}$ , the value of  $b_{A2d}$  may be shown as

$$b_{A2d} = c_{A2xyd} / c_{A2xxd}, \quad (3)$$

where

$$\begin{aligned} c_{A2xyd} &= \sum_{h=1}^H \left( \frac{1}{n_h} - \frac{1}{n_h'} \right) \left( p_{hd} s_{bxyhd} + p_{hd} q_{hd} \hat{X}_{hd} \hat{Y}_{hd} \right), \\ c_{A2xxd} &= \sum_{h=1}^H \left( \frac{1}{n_h} - \frac{1}{n_h'} \right) \left( p_{hd} s_{bxxhd}^2 + p_{hd} q_{hd} \hat{X}_{hd}^2 \right), \\ p_{hd} &= \frac{n_{hd}}{n_h}, \quad q_{hd} = 1 - p_{hd}, \\ s_{bxxhd}^2 &= \frac{1}{n_{hd} - 1} \sum_{i=1}^{n_{hd}} \left( \frac{M_{hi} p_{hid} \bar{x}_{hid}}{z_{hi}} - \hat{X}_{hd} \right)^2, \\ s_{bxyhd} &= \frac{1}{n_{hd} - 1} \sum_{i=1}^{n_{hd}} \left( \frac{M_{hi} p_{hid} \bar{x}_{hid}}{z_{hi}} - \hat{X}_{hd} \right) \left( \frac{M_{hi} p_{hid} \bar{y}_{hid}}{z_{hi}} - \hat{Y}_{hd} \right), \\ \bar{x}_{hid} &= \frac{1}{m_{hid}} \sum_{j=1}^{m_{hid}} x_{hij.d}, \quad \bar{y}_{hid} = \frac{1}{m_{hid}} \sum_{j=1}^{m_{hid}} y_{hij.d}, \\ \hat{X}_{hd} &= \frac{1}{n_{hd}} \sum_{i=1}^{n_{hd}} \frac{M_{hi} p_{hid} \bar{x}_{hid}}{z_{hi}}, \quad \hat{Y}_{hd} = \frac{1}{n_{hd}} \sum_{i=1}^{n_{hd}} \frac{M_{hi} p_{hid} \bar{y}_{hid}}{z_{hi}}. \end{aligned}$$

An approximate estimate of the variance of the linear regression estimator  $\hat{Y}_{lr2d}$  is given by

$$\hat{V}(\hat{Y}_{lr2d}) = \sum_{h=1}^H \left( \frac{1}{n_h} - r_{A2d}^2 \left( \frac{1}{n_h} - \frac{1}{n_h'} \right) \right) \left( p_{hd} s_{byhd}^2 + p_{hd} q_{hd} \hat{Y}_{hd}^2 \right), \quad (4)$$

where

$$s_{byhd}^2 = \frac{1}{n_{hd} - 1} \sum_{i=1}^{n_{hd}} \left( \frac{M_{hi} p_{hid} \bar{y}_{hid}}{z_{hi}} - \hat{Y}_{hd} \right)^2, \quad r_{A2d}^2 = \frac{c_{A2xyd}^2}{c_{A2xxd} c_{A2yyd}},$$

and  $c_{A2yyd}$  is in the same functional form as  $c_{A2xxd}$ , as defined earlier.

The estimator of the percentage Coefficient of Variation (CV) of the proposed linear regression estimator of the total area under  $d^{\text{th}}$  mixture,  $\hat{Y}_{lr2d}$ , is given by

$$\% \hat{C}V(\hat{Y}_{lr2d}) = \frac{\sqrt{\hat{V}(\hat{Y}_{lr2d})}}{\hat{Y}_{lr2d}} \times 100. \quad (5)$$

The estimator of the total area for a specific crop  $c$  based on all domains at district level is given by

$$\hat{Y}_{c2} = \sum_{d=1}^{D^*} \hat{Y}_{lr2d}, \quad (6)$$

where the sum is over all the domains containing a particular crop  $c$ ,  $d=1,2,\dots,D^*$ .

An approximate estimate of the variance of the linear regression estimator  $\hat{Y}_{c2}$  is given by

$$\hat{V}(\hat{Y}_{c2}) = \sum_{d=1}^{D^*} \hat{V}(\hat{Y}_{lr2d}).$$

Then, the estimator of the percentage CV of the proposed estimator of population total,  $\hat{Y}_{c2}$ , is given by

$$\% \hat{C}V(\hat{Y}_{c2}) = \frac{\sqrt{\hat{V}(\hat{Y}_{c2})}}{\hat{Y}_{c2}} \times 100. \quad (7)$$

### 5.2.1.2. Estimation of crop yield under mixed cropping, using the stratified two-stage two-phase sampling design framework under the area frame approach

For crop yield estimation, the sample selected by means of the area frame approach was used as the sampling frame. For this purpose, in each of the EAs sampled within a subdistrict, a list of parcels growing different mixtures was first prepared using the sample for area enumeration.

To estimate the crop area by means of the area frame approach under stratified two-stage cluster sampling, suppose that a sample of  $n_h$  EAs were selected in the  $h^{\text{th}}$  subdistrict (stratum) from the set of  $N_h$  EAs (PSUs). Then, a subsample of segments was chosen from each EA selected and all parcels of the selected segments were completely enumerated for area under a specific crop mixture. While surveying the parcels for area enumeration, a list of parcels growing different crop mixtures was prepared in each EA.

Let

$m_{hid}$  = number of parcels where the  $d^{\text{th}}$  crop mixture is grown out of the total  $m_{hi}$  parcels sampled in the  $i^{\text{th}}$  EAs of the  $h^{\text{th}}$  subdistricts,  $\forall i = 1, \dots, n_{hd}; d = 1, \dots, D; h = 1, \dots, H$ . Here, it is notable that segment-wise crop mixture lists were not prepared.

In the first phase of sampling, let

$n_h''$  = number of EAs selected by SRSWOR from the set of  $n_h$  EAs (PSU)

Within the selected EAs, for each  $d^{\text{th}}$  mixture, obtain samples of

$m_{hid}''$  = number of parcels (SSUs) identified by SRSWOR from the  $m_{hid}$  parcels chosen for recording crop area within the selected EA for each of the  $d^{\text{th}}$  mixtures ( $d = 1, 2, \dots, D$ )

Let,  $x_{hijd}$  = eye-estimated harvested yield of the crop in  $j^{\text{th}}$  parcel grown in  $d^{\text{th}}$  mixture in  $i^{\text{th}}$  EA of  $h^{\text{th}}$  subdistrict ( $j = 1, 2, \dots, m_{hid}''$ ).

In the second phase of sampling, let

$n_h'''$  = number of EAs selected by SRSWOR from  $n_h''$  selected EAs

and within each EA chosen, for each mixture  $d = 1, 2, \dots, D$ ,

$m_{hid}'''$  = number of parcels selected from  $m_{hid}''$  units in  $i^{\text{th}}$  EA of  $h^{\text{th}}$  subdistrict

In the final sample, it may be observed that of the  $n_h^m$  sample EAs,  $n_{hd}^m$  EAs grow the  $d^{th}$  mixture,  $\forall d = 1, 2, \dots, D$ .

Now, from the final  $m_{hid}^m$  parcel, estimates of a crop's harvested yields under a specific mixture may be obtained by conducting CCEs for mixed cropping. The collection of data on farmer eye estimates of crop yields is to be carried out using the questionnaire included in Annex B of the Field Test Protocol Document; the same Annex also contains a questionnaire to be used when conducting CCE questionnaires.

Let

$y_{hijd}$  = harvested yield of the crop under the  $d^{th}$  mixture from  $j^{th}$  parcel within  $i^{th}$  EA of  $h^{th}$  subdistrict

By following the double sampling approach with the information from the eye-estimated yields and the harvested yield of the sampled parcels under a specific mixture (domain), a regression estimator of the average crop yield under a specific crop mixture for a district may be obtained by

$$\bar{y}_{2lrd} = \bar{y}_{2d} + b_{2d}(\bar{x}'_{2d} - \bar{x}_{2d}) \quad (8)$$

$$\text{where } \bar{y}_{2d} = \frac{1}{H} \sum_{h=1}^H \frac{1}{n_h^m} \sum_{i=1}^{n_{hd}^m} \frac{1}{m_{hid}^m} \sum_{j=1}^{m_{hid}^m} y_{hijd},$$

$$\bar{x}_{2d} = \frac{1}{H} \sum_{h=1}^H \frac{1}{n_h^m} \sum_{i=1}^{n_{hd}^m} \frac{1}{m_{hid}^m} \sum_{j=1}^{m_{hid}^m} x_{hijd},$$

$$\bar{x}'_{2d} = \frac{1}{H} \sum_{h=1}^H \frac{1}{n_h^*} \sum_{i=1}^{n_{hd}^*} \frac{1}{m_{hid}^*} \sum_{j=1}^{m_{hid}^*} x_{hijd},$$

$$b_{2d} = \frac{c_{2xyd}}{c_{2xxd}}$$

$$c_{2xyd} = \frac{1}{H^2} \sum_{h=1}^H \left[ \left( \frac{1}{n_h^*} - \frac{1}{n_h^m} \right) (p_{hd} s_{bxyhd} + p_{hd} q_{hd} \bar{x}_{hd} \bar{y}_{hd}) + \frac{1}{n_h^* n_h^m} \sum_i \left( \frac{1}{m_{hid}^*} - \frac{1}{m_{hid}^m} \right) s_{xyhid} \right],$$

$$c_{2xxd} = \frac{1}{H^2} \sum_{h=1}^H \left[ \left( \frac{1}{n_h^m} - \frac{1}{n_h} \right) (p_{hd} s_{bxhd}^2 + p_{hd} q_{hd} \bar{x}_{hd}^2) + \frac{1}{n_h^m n_h} \sum_{i=1}^{n_{hd}^m} \left( \frac{1}{m_{hid}^m} - \frac{1}{m_{hid}} \right) s_{xhid}^2 \right],$$

$$s_{bxhd}^2 = \frac{1}{n_{hd}^m - 1} \sum_{i=1}^{n_{hd}^m} (\bar{x}_{hid} - \bar{x}_{hd})^2, \quad s_{byhd} = \frac{1}{n_{hd}^m - 1} \sum_{i=1}^{n_{hd}^m} (\bar{x}_{hid} - \bar{x}_{hd})(\bar{y}_{hid} - \bar{y}_{hd}),$$

$$s_{xhid}^2 = \frac{1}{m_{hid}^m - 1} \sum_{j=1}^{m_{hid}^m} (x_{hijd} - \bar{x}_{hid})^2,$$

$$s_{xyhid} = \frac{1}{m_{hid}^m - 1} \sum_{j=1}^{m_{hid}^m} (x_{hijd} - \bar{x}_{hid})(y_{hijd} - \bar{y}_{hid}),$$

$$p_{hd} = \frac{n_{hd}^m}{n_h^m}, \quad q_{hd} = 1 - p_{hd}, \quad p_{hid} = \frac{m_{hid}^m}{m_{hi}^m}, \quad q_{hid} = 1 - p_{hid},$$

$$\bar{x}_{hid} = \frac{1}{m_{hid}^m} \sum_j x_{hijd}, \quad \bar{y}_{hid} = \frac{1}{m_{hid}^m} \sum_j y_{hijd}, \text{ and}$$

$$\bar{x}_{hd} = \frac{1}{n_{hd}^m} \sum_{i=1}^{n_{hd}^m} \bar{x}_{hid}, \quad \bar{y}_{hd} = \frac{1}{n_{hd}^m} \sum_{i=1}^{n_{hd}^m} \bar{y}_{hid}.$$

An approximate estimate of the variance and of the percentage CV of the proposed regression estimator of the average yield of a crop under the  $d^{\text{th}}$  mixture,  $\bar{y}_{2lrd}$ , is given by

$$\hat{V}(\bar{y}_{2lrd}) = \frac{1}{H^2} \sum_{h=1}^H \left\{ \left[ \left( \frac{1}{n_h^m} - \frac{1}{N_h} \right) (p_{hd} s_{byhd}^2 + p_{hd} q_{hd} \bar{y}_{hd}^2) + \frac{1}{n_h^m N_h} \sum_{i=1}^{n_{hd}^m} \left( \frac{1}{m_{hid}^m} - \frac{1}{M_{hid}} \right) s_{yhid}^2 \right] \right. \\ \left. - r_{2d}^2 \left[ \left( \frac{1}{n_h^m} - \frac{1}{n_h} \right) (p_{hd} s_{bxhd}^2 + p_{hd} q_{hd} \bar{x}_{hd}^2) + \frac{1}{n_h^m n_h} \sum_{i=1}^{n_{hd}^m} \left( \frac{1}{m_{hid}^m} - \frac{1}{m_{hid}} \right) s_{xhid}^2 \right] \right\}, \quad (9)$$

and

$$\% \hat{CV}(\bar{y}_{2lrd}) = \frac{\sqrt{\hat{V}(\bar{y}_{2lrd})}}{\bar{y}_{2lrd}} \times 100. \quad (10)$$

where

$$s_{byhd}^2 = \frac{1}{n_{hd}^m - 1} \sum_{i=1}^{n_{hd}} (\bar{y}_{hid} - \bar{y}_{hd})^2, \quad s_{yhid}^2 = \frac{1}{m_{hid}^m - 1} \sum_{j=1}^{m_{hid}} (y_{hijd} - \bar{y}_{hid})^2,$$

$$r_{2d}^2 = \frac{c_{2xyd}^2}{c_{2xxd} c_{2yyd}},$$

and  $c_{2yyd}$  is in the same functional form as  $c_{2xxd}$  (defined earlier).

The estimator of crop yield for a specific crop  $c$  based on all domains at district level, and its approximate estimator of variance, is given by

$$\hat{Y}_{c2} = \frac{1}{D^*} \sum_{d=1}^{D^*} \bar{y}_{2lrd} \quad (11)$$

and

$$\hat{V}(\hat{Y}_{c2}) = \frac{1}{D^{*2}} \sum_{d=1}^{D^*} \hat{V}(\bar{y}_{2lrd}), \quad (12)$$

where the sum includes all domains containing a particular crop  $c$  in the different mixtures,  $d=1,2,\dots,D^*$ .

Then, the estimator of the percentage CV of the proposed estimator of population total,  $\hat{Y}_{c2}$ , is given by

$$\% \hat{CV}(\hat{Y}_{c2}) = \frac{\sqrt{\hat{V}(\hat{Y}_{c2})}}{\hat{Y}_{c2}} \times 100. \quad (13)$$

## 5.2.2. The household approach

List frames are most appropriate when households are to be sampled, and are the most common type of frames in agricultural probability sample surveys. List frames are developed on the basis of information collected from the most recent agricultural or population census, administrative data, previous surveys or a combination of several data sources. If their component units are clusters, it may be necessary to create multistage sampling schemes with further frames to refine the identification of the clusters' elements and thus reach the target population.

Auxiliary information may be available in list frames, which would enable the use of efficient sampling schemes such as stratified sampling, probability-proportional-to-size sampling or even both of these techniques simultaneously, as well as use of calibration and regression-type estimators. The sampling and identification of reporting units in agricultural surveys have a relatively low cost when list frames are used, because the sampled farmers' names and addresses are listed either as the one-stage or multistage final sampling units, and are thus promptly accessible for the fieldwork. However, list frames become obsolete quickly, resulting in problems of undercoverage and obsolete information. This emphasizes the need to properly maintain list frames, to avoid problems of duplicate records from arising.

If the parcels in the selected villages or EAs cannot be identified due to the unavailability of land cadastral maps\* and parcel-wise crop registers, it is advisable to adopt the household approach. Therefore, villages, EAs or Census Blocks (CBs) may be considered as the primary stage unit and a sample of households may be selected from each unit selected during the first stage. To estimate the crop area, all the fields belonging to a selected household may be enumerated.

For crop area estimation, a stratified two-stage sampling design with two phases within each stage was adopted.

---

\* Land cadastral maps are a set of records showing the extent

value and ownership (or other basis for use or occupancy) of land. Technically a cadastre is a record of areas and values of land and of landholders that was originally compiled for taxation purposes. However several countries no longer impose any land tax and the cadastre serves two other equally important purpose: (1) it provides a ready means of precise description and identification of particular pieces of land and (2) it acts as a continuous record of rights in land.

---

For further details, see “Cadastral Maps”, in Bins, B.O. (Dale, P.F. rev.), *Cadastral surveys and records of rights in land*, FAO Publication: Rome. Available at <http://www.fao.org/docrep/006/v4860e/v4860e03.htm>. Accessed 9 February 2017.

### 5.2.2.1. Estimation procedure for crop area estimation under mixed cropping using the household approach

Let

$H$  = number of subdistricts in a district considered as  $H$  strata

$N_h$  = number of Enumeration Districts (EDs) or CBs in the  $h$ th subdistrict (stratum) considered as the PSUs ( $h=1, \dots, H$ ). Generally, the total number of villages or CBs in each subdistrict,  $N_h$ , is known

$M_{hi}$  = number of households (SSUs) in  $i$ th ED or CB of  $h$ th stratum

$T_{hij}$  = number of parcels (USUs) in  $j$ th household of  $i$ th ED or CB where different forms of crop mixture (e.g. pure stand, mixture-1 and mixture-2) are grown

The different crop mixtures are taken as the domains of the study. It is assumed that in each subdistrict  $h$ ,  $D$  different crop mixtures are followed as pure stand, mixture-1, mixture-2, etc. In this case, there would be  $\{U_{h1}, \dots, U_{hd}, \dots, U_{hD}\}$  domains in the  $h$ th stratum,  $U_h$ . Although the number of villages or CBs ( $N_h$ ), the number of households ( $M_{hi}$ ) and the number of parcels ( $T_{hij}$ ) in each subdistrict are known, the number of villages or CBs ( $N_{hd}$ ), the number of households ( $M_{hid}$ ) and the number of parcels ( $T_{hijd}$ ) for a particular mixture are generally unknown.

Let

$y_{hijk}$  = crop area of  $k$ th parcel (USU) in  $j$ th household (SSU) of  $i$ th ED or CB (PSU) in the  $h$ th subdistrict (stratum) of a district

The total area under the  $d$ th crop mixture (domain) in a district is given by

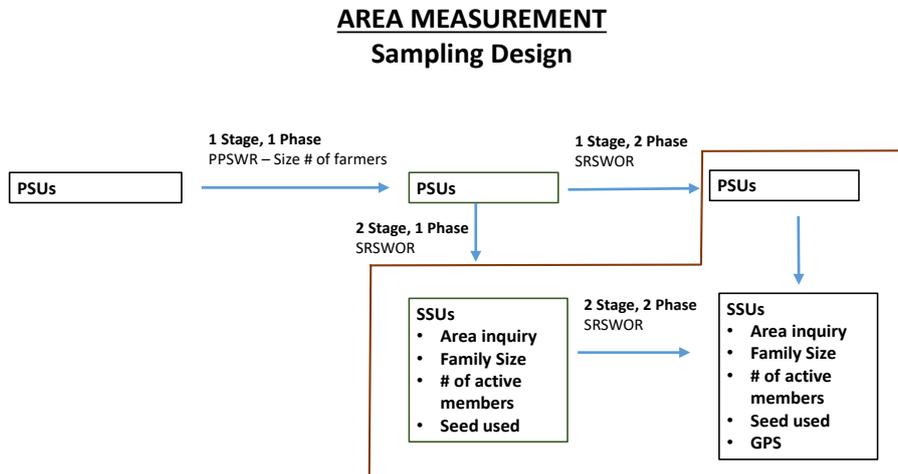
$$Y_d = \sum_{h=1}^H \sum_{i=1}^{N_{hd}} \sum_{j=1}^{M_{hid}} \sum_{k=1}^{T_{hijd}} y_{hijk} \quad (14)$$

The population total based on all domains is given as

$$Y = \sum_{d=1}^D Y_d = \sum_{d=1}^D \sum_{h=1}^H \sum_{i=1}^{N_{hd}} \sum_{j=1}^{M_{hid}} \sum_{k=1}^{T_{hijd}} y_{hijk}$$

The proposed sampling design for estimating crop area at the district level using the household approach is the stratified two-stage cluster sampling design, with two phases to each stage of sampling.

Figure 5.2.2.1.1. Sampling design for area measurement



Let, in the first stage,

$n'_h$  = number of EDs or CBs drawn from  $N_h$  EDs or CBs (PSUs) using PPSWR

The probability of selecting the  $i^{th}$  ED or CB in the  $h^{th}$  stratum is computed as  $Z_{hi}' = X_{hi}/X_h$ , where  $X$  may be considered as the total number of agricultural households.

In the second phase of the first stage,

$m_{hi}'$  = number of households selected by SRSWOR; all  $T_{hij}$  parcels in the selected household are completely enumerated to collect auxiliary information on the parcel, such as on the seed used or farmer assessments

$n_{hd}'$  = number of EDs or CBs of the  $n_h'$  EDs or CBs selected that grow the specific  $d^{th}$  mixture.

$m_{hid}'$  = number of households, of the  $m_{hi}'$  households selected, that growing the  $d^{th}$  mixture in the  $T_{hijd}$  parcels

In the second stage,

$n_h$  = number of EDs or CBs selected from the  $n_h'$  ED or CB (PSU) initially selected by SRSWOR

In each ED or CB selected, a subsample of:

$m_{hi}$  = number of households selected by SRSWOR from  $m_{hi}'$  households; all  $T_{hij}$  parcels in the selected household are completely enumerated. Within these sampled households, the areas of each parcel where crop mixtures are grown are measured by GPS. Data was collected using the questionnaires provided in Annex C of the Field Test Protocol document. The area where component crops are grown was obtained by apportioning, using the information on seed rates.

$n_{hd}$  = number of EDs or CBs of  $n_h$  EDs or CBs growing  $d^{th}$  mixture.

$m_{hid}$  = number of households of  $m_{hi}$  households growing  $d^{th}$  mixture in  $T_{hijd}$  parcels

In the current scenario, the aim is to estimate the total crop area under a specific crop ( $Y$ ) and under different mixtures ( $Y_d$ ), where  $d = 1, 2, \dots, D$ .

Let

$x_{hijk}$  = auxiliary information (e.g. seed used or farmer assessment) corresponding to  $k^{\text{th}}$  parcel (SSU) of  $j^{\text{th}}$  selected household in  $i^{\text{th}}$  ED or CB (PSU) within  $h^{\text{th}}$  subdistrict (stratum)

$y_{hijk}$  = crop area measured by GPS of  $k^{\text{th}}$  parcel (USU) in  $j^{\text{th}}$  household (SSU) of  $i^{\text{th}}$  ED or CB (PSU) in  $h^{\text{th}}$  subdistrict (stratum) of a district

A double sampling regression estimator of the total area growing the  $d^{\text{th}}$  mixture under the stratified two-phase two-stage cluster sampling design may be computed as

$$\hat{Y}_{lr3d} = \hat{Y}_d + b_{A3d} (\hat{X}'_d - \hat{X}_d), \quad (15)$$

where

$$\begin{aligned} \hat{Y}_d &= \sum_{h=1}^H \frac{1}{n_h} \sum_{i=1}^{n_{hd}} \frac{1}{z'_{hi}} \frac{M_{hi}}{m_{hi}} \sum_{j=1}^{m_{hid}} \sum_{k=1}^{T_{hijd}} y_{hijk} = \sum_{h=1}^H \frac{1}{n_h} \sum_{i=1}^{n_{hd}} \frac{1}{z'_{hi}} \frac{M_{hi}}{m_{hi}} \sum_{j=1}^{m_{hid}} y_{hij} = \sum_{h=1}^H \frac{1}{n_h} \sum_{i=1}^{n_{hd}} \frac{\hat{Y}_{hid}}{z'_{hi}}, \\ \hat{X}_d &= \sum_{h=1}^H \frac{1}{n_h} \sum_{i=1}^{n_{hd}} \frac{\hat{X}_{hid}}{z'_{hi}}, \quad \hat{X}'_d = \sum_{h=1}^H \frac{1}{n_h} \sum_{i=1}^{n_{hd}} \frac{\hat{X}'_{hid}}{z'_{hi}}, \\ \hat{Y}_{hid} &= \frac{M_{hi}}{m_{hi}} \sum_{j=1}^{m_{hid}} \sum_{k=1}^{T_{hijd}} y_{hijk}, \quad \hat{X}_{hid} = \frac{M_{hi}}{m_{hi}} \sum_{j=1}^{m_{hid}} \sum_{k=1}^{T_{hijd}} x_{hijk}, \quad \hat{X}'_{hid} = \frac{M_{hi}}{m'_{hi}} \sum_{j=1}^{m'_{hid}} \sum_{k=1}^{T_{hijd}} x_{hijk}. \end{aligned}$$

By minimizing the variance of the double sampling regression estimator  $\hat{Y}_{lr3d}$  with respect to  $b_{A3d}$  and ignoring the variation in  $b_{A3d}$ , the value of  $b_{A3d}$  may be shown as

$$b_{A3d} = c_{A3xyd} / c_{A3xxd}, \quad (16)$$

where

$$\begin{aligned}
c_{A3xyd} &= \sum_{h=1}^H \left( \frac{1}{n_h} - \frac{1}{n_h'} \right) \left( p_{hd} s_{bxyhd} + p_{hd} q_{hd} \hat{X}_{hd} \hat{Y}_{hd} \right), \\
c_{A3xxd} &= \sum_{h=1}^H \left( \frac{1}{n_h} - \frac{1}{n_h'} \right) \left( p_{hd} s_{bxxhd}^2 + p_{hd} q_{hd} \hat{X}_{hd}^2 \right), \\
p_{hd} &= \frac{n_{hd}}{n_h}, \quad q_{hd} = 1 - p_{hd}, \\
s_{bxxhd}^2 &= \frac{1}{n_{hd} - 1} \sum_{i=1}^{n_{hd}} \left( \frac{M_{hi} p_{hid} \bar{x}_{hid}}{z_{hi}'} - \hat{X}_{hd} \right)^2, \\
s_{bxyhd} &= \frac{1}{n_{hd} - 1} \sum_{i=1}^{n_{hd}} \left( \frac{M_{hi} p_{hid} \bar{x}_{hid}}{z_{hi}'} - \hat{X}_{hd} \right) \left( \frac{M_{hi} p_{hid} \bar{y}_{hid}}{z_{hi}'} - \hat{Y}_{hd} \right), \\
\bar{x}_{hid} &= \frac{1}{m_{hid}} \sum_{j=1}^{m_{hid}} x_{hij.d}, \quad \bar{y}_{hid} = \frac{1}{m_{hid}} \sum_{j=1}^{m_{hid}} y_{hij.d}, \\
\hat{X}_{hd} &= \frac{1}{n_{hd}} \sum_{i=1}^{n_{hd}} \frac{M_{hi} p_{hid} \bar{x}_{hid}}{z_{hi}'}, \quad \hat{Y}_{hd} = \frac{1}{n_{hd}} \sum_{i=1}^{n_{hd}} \frac{M_{hi} p_{hid} \bar{y}_{hid}}{z_{hi}'}.
\end{aligned}$$

An approximate estimate of the variance of the double sampling regression estimator  $\hat{Y}_{lr3d}$  is given by

$$\hat{V}(\hat{Y}_{lr2d}) = \sum_{h=1}^H \left( \frac{1}{n_h} - r_{A2d}^2 \left( \frac{1}{n_h} - \frac{1}{n_h'} \right) \right) \left( p_{hd} s_{byhd}^2 + p_{hd} q_{hd} \hat{Y}_{hd}^2 \right), \tag{17}$$

where

$$s_{byhd}^2 = \frac{1}{n_{hd} - 1} \sum_{i=1}^{n_{hd}} \left( \frac{M_{hi} p_{hid} \bar{y}_{hid}}{z_{hi}'} - \hat{Y}_{hd} \right)^2, \quad r_{A3d}^2 = \frac{c_{A3xyd}^2}{c_{A3xxd} c_{A3yyd}},$$

and  $c_{A3yyd}$  is in same functional form as  $c_{A3xxd}$  (as defined earlier).

Then, the estimator of the percentage CV of the proposed linear regression estimator of the total area under the  $d$ th mixture,  $\hat{Y}_{lr3d}$ , is given by

$$\% \hat{C}V(\hat{Y}_{lr3d}) = \frac{\sqrt{\hat{V}(\hat{Y}_{lr3d})}}{\hat{Y}_{lr3d}} \times 100. \quad (18)$$

The estimator of the total area for a specific crop  $c$  based on all the domains at district level is given by

$$\hat{Y}_{c3} = \sum_{d=1}^{D^*} \hat{Y}_{lr3d}, \quad (19)$$

where the sum is over all those domains containing a particular crop  $c$ ,  $d=1,2,\dots,D^*$ .

An approximate estimate of variance of the double sampling regression estimator  $\hat{Y}_{c3}$  is given by

$$\hat{V}(\hat{Y}_{c3}) = \sum_{d=1}^{D^*} \hat{V}(\hat{Y}_{lr3d}) \quad (20)$$

Then, the estimator of the percentage CV of the proposed estimator of population total  $\hat{Y}_{c3}$  is given by

$$\% \hat{C}V(\hat{Y}_{c3}) = \frac{\sqrt{\hat{V}(\hat{Y}_{c3})}}{\hat{Y}_{c3}} \times 100. \quad (21)$$

#### 5.2.2.2. Estimation of crop yield for mixed cropping using a stratified two-stage two-phase sampling design

The **household approach** is appropriate for crop yield estimation when the records do not contain any information on the parcels and the EDs or CBs have distinct boundaries. The sample selected for crop area estimation provides the sampling frame for drawing the sample for crop yield estimation. Accordingly,

a list of parcels growing different mixtures was prepared using the results of area enumeration conducted in each ED or CB sampled in a subdistrict.

For “Estimation of crop area” by means of the household approach under the stratified two-phase two-stage cluster sampling design, suppose that in the  $h^{\text{th}}$  subdistrict (stratum) of the set of  $N_h$  EDs or CBs (PSUs), a sample of  $n_h$  EDs or CBs are selected. Then, a subsample of households is drawn from each selected ED or CB and all the parcels of the households selected are completely enumerated for area under a specific crop mixture. While the parcels are surveyed for area enumeration, a list of parcels growing different mixtures is prepared for a given ED or CB.

Let

$m_{hid}$  = number of parcels grown as the  $d^{\text{th}}$  mixture of the crop, out of the  $m_{hi}$  sampled parcels in the  $n_{hd}$  EDs or CBs of the  $h^{\text{th}}$  subdistricts,  $\forall d = 1, 2, \dots, D$

In the first phase of sampling,

$n_h''$  = number of EDs or CBs selected by SRSWOR from the set of  $n_h$  EDs or CBs (PSUs). Within the EDs or CBs selected, for each of the  $d^{\text{th}}$  mixture,  $d = 1, 2, \dots, D$ , let

$m_{hid}''$  = number of parcels (SSUs) selected by SRSWOR from the  $m_{hid}$  parcels, used for recording area, within the selected ED or CB for each of the  $d^{\text{th}}$  mixture,  $d = 1, 2, \dots, D$

For the selected parcels, farmer eye estimates of the harvested yield of the crop under mixture were recorded.

$x_{hijd}$  = eye-estimated harvested yield of crop in  $j^{\text{th}}$  parcel grown under  $d^{\text{th}}$  mixture in  $i^{\text{th}}$  ED or CB of  $h^{\text{th}}$  subdistrict

In the second phase of sampling,

$n_h'''$  = number of EDs or CBs selected by SRSWOR from the  $n_h''$  EDs or CBs selected

$m_{hid}'''$  = number of parcels selected from the  $m_{hid}''$  first-stage, second-phase units in  $i^{\text{th}}$  ED or CB of  $h^{\text{th}}$  subdistrict for each mixture,  $d=1, 2, \dots, D$ .

In the final sample, it may be observed that of the  $n_h'''$  sample EDs or CBs,  $n_{hd}'''$  EDs or CBs follow the  $d^{\text{th}}$  mixture,  $\forall d = 1, 2, \dots, D$ . For all  $m_{hid}'''$  parcels in the final sample, estimates of the harvested yields of a crop under a specific mixture are obtained by conducting CCEs or similar processes for mixed cropping. Data on farmer eye estimates of crop yields is to be collected using Questionnaire-6 (C-3) in Annex C of the Field Test Protocol Document; for data collection on crop yield, the CCE questionnaires provided in the same Annex are to be used.

Let

$y_{hijd}$  = harvested yield of the crop under  $d^{\text{th}}$  mixture from  $j^{\text{th}}$  parcel in  $i^{\text{th}}$  ED or CB of  $h^{\text{th}}$  subdistrict

Utilizing the information from the eye-estimated yields and the harvested yield of the sampled parcels under a specific mixture (domain), and using double sampling approach, a double-sampling regression estimator of the average crop yield under a specific crop mixture for a given district may be obtained by

$$\bar{y}_{lr3d} = \bar{y}_{3d} + b_{3d}(\bar{x}'_{3d} - \bar{x}_{3d}), \quad (22)$$

$$\bar{y}_{3d} = \frac{1}{H} \sum_{h=1}^H \frac{1}{n_h} \sum_{i=1}^{n_{hd}} \frac{1}{m_{hid}} \sum_{j=1}^{m_{hid}} y_{hijd},$$

where

$$\bar{x}_{3d} = \frac{1}{H} \sum_{h=1}^H \frac{1}{n_h} \sum_{i=1}^{n_{hd}} \frac{1}{m_{hid}} \sum_{j=1}^{m_{hid}} x_{hijd},$$

$$\bar{x}'_{3d} = \frac{1}{H} \sum_{h=1}^H \frac{1}{n_h} \sum_{i=1}^{n_{hd}} \frac{1}{m_{hid}} \sum_{j=1}^{m_{hid}} x'_{hijd},$$

$$b_{3d} = \frac{c_{3xyd}}{c_{3xxd}}$$

$$c_{3xyd} = \frac{1}{H^2} \sum_{h=1}^H \left[ \left( \frac{1}{n_h} - \frac{1}{n_h} \right) (p_{hd} S_{bxyhd} + p_{hd} q_{hd} \bar{x}_{hd} \bar{y}_{hd}) + \frac{1}{n_h n_h} \sum_i^{n_{hd}} \left( \frac{1}{m_{hid}} - \frac{1}{m_{hid}} \right) S_{xyhid} \right],$$

$$c_{3xxd} = \frac{1}{H^2} \sum_{h=1}^H \left[ \left( \frac{1}{n_h} - \frac{1}{n_h} \right) (p_{hd} S_{bxxhd}^2 + p_{hd} q_{hd} \bar{x}_{hd}^2) + \frac{1}{n_h n_h} \sum_i^{n_{hd}} \left( \frac{1}{m_{hid}} - \frac{1}{m_{hid}} \right) S_{xxhid}^2 \right],$$

$$S_{bxxhd}^2 = \frac{1}{n_{hd} - 1} \sum_{i=1}^{n_{hd}} (\bar{x}_{hid} - \bar{x}_{hd})^2, \quad S_{bxyhd} = \frac{1}{n_{hd} - 1} \sum_{i=1}^{n_{hd}} (\bar{x}_{hid} - \bar{x}_{hd})(\bar{y}_{hid} - \bar{y}_{hd}),$$

$$S_{xhid}^2 = \frac{1}{m_{hid}^m - 1} \sum_{j=1}^{m_{hid}^m} (x_{hijd} - \bar{x}_{hid})^2, \quad S_{xyhid} = \frac{1}{m_{hid}^m - 1} \sum_{j=1}^{m_{hid}^m} (x_{hijd} - \bar{x}_{hid})(y_{hijd} - \bar{y}_{hid}),$$

$$p_{hd} = \frac{n_{hd}^m}{n_h^m}, \quad q_{hd} = 1 - p_{hd}, \quad p_{hid} = \frac{m_{hid}^m}{m_{hi}^m}, \quad q_{hid} = 1 - p_{hid},$$

$$\bar{x}_{hid} = \frac{1}{m_{hid}^m} \sum_j x_{hijd}, \quad \bar{y}_{hid} = \frac{1}{m_{hid}^m} \sum_j y_{hijd},$$

$$\bar{x}_{hd} = \frac{1}{n_{hd}^m} \sum_{i=1}^{n_{hd}^m} \bar{x}_{hid}, \quad \bar{y}_{hd} = \frac{1}{n_{hd}^m} \sum_{i=1}^{n_{hd}^m} \bar{y}_{hid}.$$

An approximate estimate of the variance and percentage CV of the proposed double-sampling regression estimator of the average yield of a crop under the  $d^{\text{th}}$  mixture,  $\bar{y}_{lr3d}$ , is given by

and

$$\% CV(\bar{y}_{lr3d}) = \frac{\sqrt{\hat{V}(\bar{y}_{lr3d})}}{\bar{y}_{lr3d}} \times 100, \quad (24)$$

where

$$S_{byhd}^2 = \frac{1}{n_{hd}^m - 1} \sum_{i=1}^{n_{hd}^m} (\bar{y}_{hid} - \bar{y}_{hd})^2, \quad S_{yhid}^2 = \frac{1}{m_{hid}^m - 1} \sum_{j=1}^{m_{hid}^m} (y_{hijd} - \bar{y}_{hid})^2,$$

$$r_{3d}^2 = \frac{c_{3xyd}^2}{c_{3xxd} c_{3yyd}},$$

and  $c_{3yyd}$  is in the same functional form as  $c_{3xxd}$  (defined earlier).

The estimator of yield for a specific crop  $c$  based on all domains at district level, together with its approximate estimator of variance, is given by

$$\hat{Y}_{c3} = \frac{1}{D^*} \sum_{d=1}^{D^*} \bar{y}_{lr3d} \quad (25)$$

and

$$\hat{V}\left(\hat{Y}_{c3}\right) = \frac{1}{D^{*2}} \sum_{d=1}^{D^*} \hat{V}\left(\bar{y}_{lr3d}\right), \quad (26)$$

where the sum includes all those domains containing a particular crop  $c$  in the different mixtures,  $d=1, 2, \dots, D^*$ .

Then, the estimator of the percentage CV of the proposed estimator of population total  $\hat{Y}_{c3}$  is given by

$$\% \hat{CV}\left(\hat{Y}_{c3}\right) = \frac{\sqrt{\hat{V}\left(\hat{Y}_{c3}\right)}}{\hat{Y}_{c3}} \times 100. \quad (27)$$

For further information on the implementation of this method and data collection processes, with examples from three countries, see Sud et al. (2015).

# Results of Three Country Field Tests

## 6.1. Introduction

The sampling and estimation methodology described in Chapter 5, and some of the data collection methods examined in Chapter 2-4, were piloted in Indonesia, Jamaica and Rwanda. In particular, the household approach was applied in Indonesia and Jamaica, while the area frame approach was applied in Rwanda. The fieldwork was implemented by the *Badan Pusat Statistik* (BPS) of Indonesia, the Ministry of Industry, Commerce, Agriculture and fisheries (MICAF) of Jamaica, and the National Institute of Statistics of Rwanda for the three countries, respectively. This chapter provides the results and draws comparisons between the methods piloted.

## 6.2. Results of the field tests

### 6.2.1. Indonesia

#### Crop area estimation

In Indonesia, crop area measured by means of GPS was considered as the main variable, while crop area by farmer inquiry, the active family members and the total number of family members were considered auxiliary variables. Figure 6.2.1.1 presents a scatter plot between the GPS measurements and the area measurements obtained by means of farmer inquiry. It can be seen that there is a near-linear relationship between the area of the parcel as measured by GPS and that determined by reference to farmer inquiry; in addition, the magnitude of  $R^2$  is shown to be equal to 0.95.

**Figure 6.2.1.1. Scatter plot indicating parcel area by GPS and farmer inquiry**

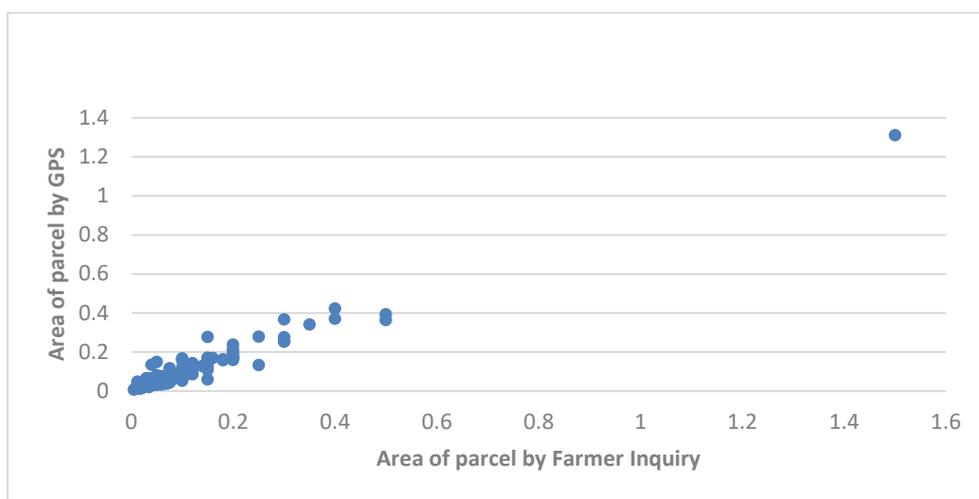


Table 6.2.1.1 provides an unweighted correlation matrix indicating the correlation between crop area by GPS, crop area by Inquiry, the active family members and the total number of family members. The table shows that the correlation between crop area illustrated by GPS and crop area determined by inquiry is very high. However, the correlation between the active family members and the total number of family members with the crop area (both by GPS and by inquiry) is poor. However, active family members exhibit a high correlation with the total number of family members.

**Table 6.2.1.1. Unweighted correlation matrix indicating the correlation between crop area by GPS, crop area by inquiry, active family members and total number of family members**

<b>Correlation</b>	<b>GPS</b>	<b>Inquiry</b>	<b>Active members</b>	<b>Total members</b>
GPS	1.00	0.98	0.23	0.13
Inquiry	0.98	1.00	0.22	0.13
Active members	0.23	0.22	1.00	0.70
Total members	0.13	0.13	0.70	1.00

Table 6.2.1.2 provides the area-classwise average difference and average relative difference between the parcel area as measured by GPS and that ascertained by means of farmer inquiry. It is evident from the results that the average relative difference between the parcel area established by farmer inquiry and that determined by GPS declines as the parcel size increases.

**Table 6.2.1.2. Area-class-wise average difference and average relative difference between parcel area, as measured by GPS and by farmer inquiry**

<b>Area class</b>	<b>Number of parcels</b>	<b>Average difference between parcel area measured by GPS and by farmer inquiry</b>	<b>Average percentage relative difference between parcel area measured by GPS and by farmer inquiry</b>
<= 0.1 ha	84	0.01	11.64
> 0.1 ha & <= 0.2 ha	24	0.00	-2.30
> 0.2 ha & <= 0.3 ha	7	-0.01	-4.91
> 0.3 ha	6	0.03	8.46
Overall	121	0.00	7.76

Table 6.2.1.3 provides crop-wise estimates of crop area by GPS, GPS using crop area by inquiry, GPS and active family members, and GPS and total number of family members as auxiliary variables. The table shows that generally, the percentage CV of an estimator that uses farmer-reported crop area as an auxiliary variable is minimal. An estimator that does not use auxiliary information exhibits a high percentage CV. It may also be seen that the percentage CVs of all of the estimators are higher than the prescribed percentage CV for estimation at the district level. This is attributable to the fact that for this study, small sample sizes were used in Indonesia.

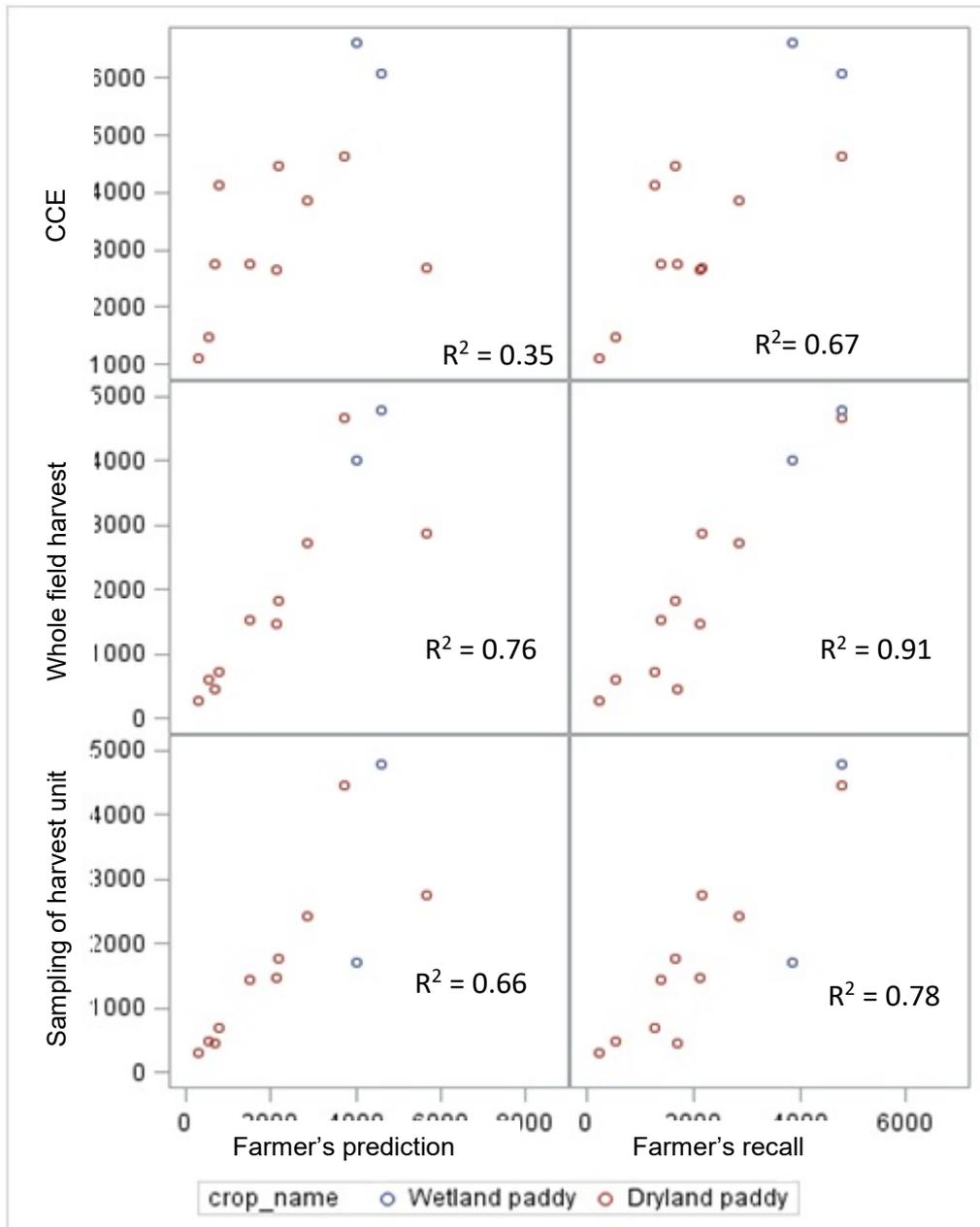
**Table 6.2.1.3. Weighted crop-wise regression estimator of crop area by GPS, using crop area by inquiry, active family members and total number of family members as auxiliary variables**

Crop name	Crop-wise sample sizes	Crop area estimates by GPS only		Crop area estimates using total number of family members as auxiliary variables with %CV		Crop area estimates using active family members as auxiliary variables, with %CV		Crop area estimates using crop area by inquiry as auxiliary variables, with %CV	
		Crop area estimates (ha)	%CV	Crop area estimates (ha)	%CV	Crop area estimates (ha)	%CV	Crop area estimates (ha)	%CV
Dryland paddy	81	548.74	22.18	532.26	21.13	512.91	21.72	599.98	18.25
Maize	41	180.08	21.59	180.32	19.40	183.93	19.03	262.42	13.26
Wetland paddy	17	125.66	47.67	109.09	48.54	124.89	41.82	174.00	30.02
Peanut	9	31.51	63.02	70.06	25.90	91.51	19.90	71.58	25.33
Soybean	2	3.23	17.89	1.35	39.19	1.35	39.19	1.35	39.19

Apportioning the crop mixture area among the component crops of the crop mixture was done using the seed used and the standard seed rate for normal crops (pure crops) provided by the BPS. The data on CCEs was available for only 12 parcels of different crops, including two parcels of two different crop mixtures. CCE data for the two mixtures was available for only one of the component crops. The district-level estimates were obtained using the apportioned area of the crop in the mixture and the area of the particular crop in the pure stand.

Figure 6.2.1.2 presents scatter plots between crop yield by farmer's prediction, farmer's recall, CCE, whole- field harvest and sampling of harvest units.

**Figure 6.2.1.2. Scatter plots between the main variables and the auxiliary variables of crop yield measurements, i.e. crop yield by farmer's prediction, farmer's recall, CCE, whole field harvest and sampling of harvest units.**



## Crop yield estimation

Table 6.2.1.4 provides estimates of crop yield together with percentage CVs using double-sampling regression estimators that involve different variables. The table shows that the method of sampling the harvest unit consistently underestimates crop production; the magnitude of such underestimation varies from 4.5 per cent to 28.7 per cent, while the CCE technique overestimates crop production. The magnitude of overestimation ranges between 46.1 to 178.1 per cent.

Estimates of dryland paddy yield were obtained using CCE yield data and by sampling the harvest units as the main variables; farmer crop predictions and recall of produce were taken as auxiliary variables. Accordingly, it was possible to obtain regression estimates based on the double-sampling approach, together with their percentage CVs. It may be seen that the magnitude of percentage CVs was lowest for estimators in which the study variable was the sampling of harvest units and the auxiliary variable was the farmer prediction of crop produce. Thus, by taking into account both the closeness of a measurement method to the whole field harvest as well as the criterion of percentage CV, it is recommended to adopt the method of sampling the harvest unit, together with farmer prediction of crop produce as an auxiliary variable. However, these estimates are based on a limited number of observations. Therefore, the recommended method should be tested on a larger scale before it is adopted. Average yield estimates were obtained by averaging the average yield obtained from pure stand and the average yield obtained from the mixture of the particular crop.

**Table 6.2.1.4. Estimates of crop yield and percentage CVs, using double-sampling regression estimators involving different variables**

Estimators	Auxiliary variable	Estimates (kg/ha)	%CV
Crop yield by CCE	Farmer prediction of crop produce	3,250	25.84
Crop yield by CCE	Farmer recall of crop produce	2,390	34.96
Crop yield by sampling of harvest units	Farmer prediction of crop produce	2,087	24.14
Crop yield by sampling of harvest units	Farmer recall of crop produce	1,327	38.36

## 6.2.2. Rwanda

### Crop area estimation

In Rwanda, crop area measured by means of GPS was considered as the main variable, whereas the auxiliary variables were the crop area established by reference to maps, the family members and the active number of family members.

Figure 6.2.2.1 presents a scatter plot between the areas of parcels as measured by GPS and that determined through maps. The figure shows that the areas of parcels measured by GPS and by maps display an almost linear relation; the magnitude of  $R^2$  is 0.898.

**Figure 6.2.2.1. Scatter plot indicating areas of parcels determined by maps and by GP**

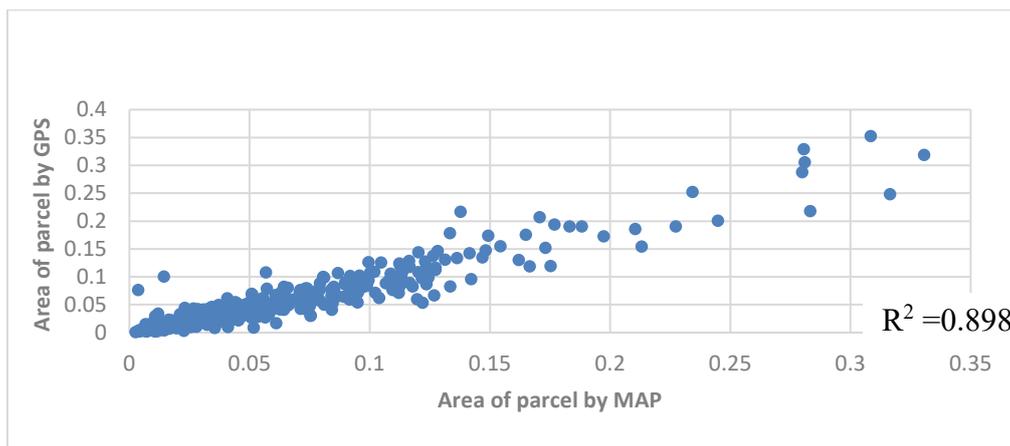


Table 6.2.2.1 shows the unweighted correlation matrix between crop area determined by map, crop area determined by GPS, the number of family members of the farmer and the farmer's active family members. From the data, it is clear that the GPS and map areas are highly correlated. However, the variable numbers of family members and of active family members exhibit poor correlation with each other, as well as with the data on the crop areas computed by GPS and by map.

**Table 6.2.2.1. Unweighted correlation matrix indicating correlation between crop area determined by map, crop area by GPS, number of family members and active family members.**

	MAP	GPS	Family member	Active family member
MAP	1.00	0.95	0.08	-0.02
GPS	0.95	1.00	0.09	-0.02
Family member	0.08	0.09	1.00	0.92
Active family member	-0.02	-0.02	0.92	1.00

**Table 6.2.2.2. Area-class-wise average relative difference between parcel areas, as measured by GPS and by maps**

Area class	Number of parcels	Average difference between maps and GPS	Average relative difference between maps and GPS
<= 0.1 ha	518	0.007865	40.6004
>0.1 & <=0.2 ha	46	0.000338	-0.1167
>0.2 & <= 0.3 ha	7	0.005795	2.484
>0.3 ha	4	-0.02582	-7.746
All	575	0.007003	36.5427

It is evident from the results shown in Table 6.2.2.2 that the average relative difference between the map-determined area and the GPS-established area declines as the parcel size increases.

Table 6.2.2.3 provides estimators based on the double-sampling approach and simple linear estimators of crop area, together with their percentage CVs. It may be seen that the regression estimator having GPS-based crop area as the dependent variable and map-based crop area as auxiliary variable is the most reliable. The performance of these estimators deteriorates when GPS crop area is used as a dependent variable and family members, active number of family members are used as auxiliary variables. Surprisingly, the regression estimators in the two cases perform poorly compared to estimators constructed taking into account only the GPS-based crop area and the enumerator's self-reported crop area. These results may be explained by the small sample size observed in the survey and the relatively low quality of the data used. The small sample size also explains the high percentage CVs of all estimates, which these vary from moderately high to very high.

**Table 6.2.2.3. Estimators based on the double-sampling approach and simple linear estimators of crop area, with their percentage CVs.**

Crop name	Crop area estimates along with %CV				Crop area regression estimates along with %CV					
	By GPS only		By map only		Using map as auxiliary variable		Using total family members as auxiliary variable		Using active members as auxiliary variable	
	Crop area estimates (ha)	% CV	Crop area estimates (ha)	% CV	Crop area estimates (GPS) (ha)	% CV	Crop area estimates (GPS) (ha)	% CV	Crop area estimates (GPS) (ha)	% CV
Climbing beans	4,790.81	22.59	5,560.33	22.35	4,125.51	17.2	4,173.76	25.76	4,302.52	24.98
Maize	13,169.02	24.7	15,652.88	23.64	11,182.26	23.83	11,954.39	27.15	12,019.59	27
Sweet potatoes	168.95	63.32	207.35	55.16	144.33	47.59	148.17	71.95	144.51	73.76

### Crop yield estimation

Figure 6.2.2.2 presents scatter plots between crop yield by farmer's prediction, farmer's recall, CCE, whole field harvest and the enumerator's self-assessment of produce. The table shows that there is a near-linear relationship between the area of the parcel measured by GPS and that determined by means of farmer inquiry.

**Figure 6.2.2.2. Scatter plots between crop yield by farmer’s prediction, farmer’s recall, CCE, whole field harvest and enumerator’s self-assessment of produce.**

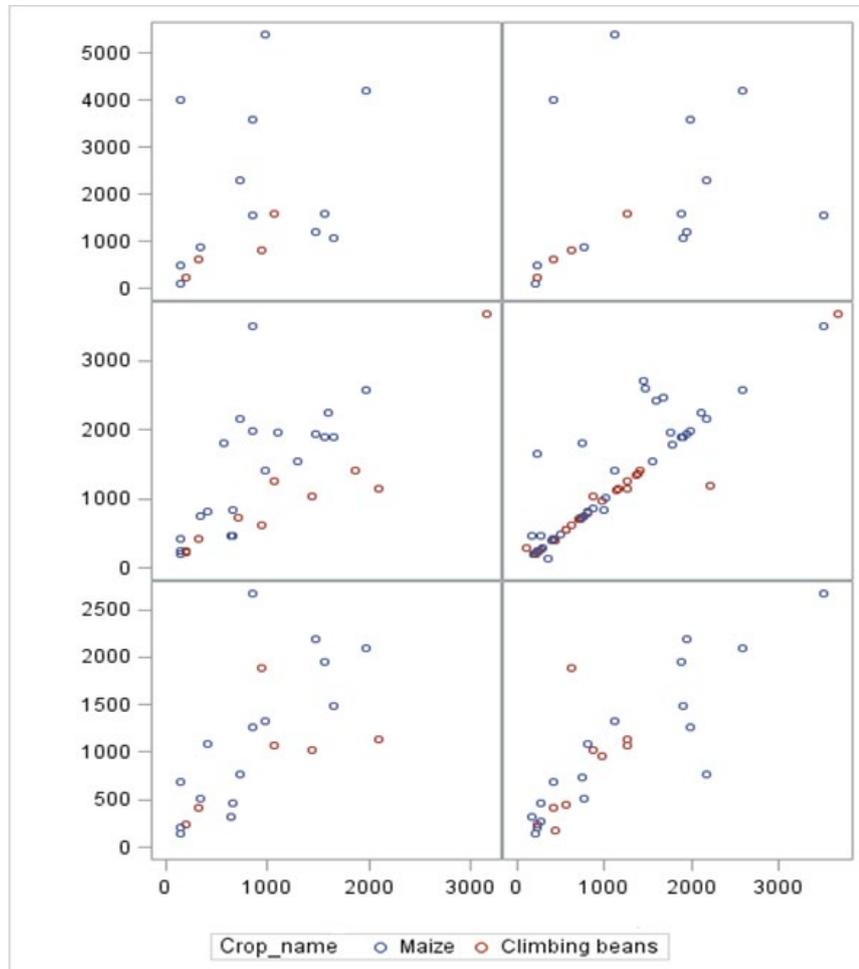


Table 6.2.2.4 provides estimates of crop yield through enumerator assessments of crop produce and CCE. For this purpose, regression estimators based on both a simple- and a double-sampling approach estimators are considered, and farmer-predicted yield and crop yield through farmer recall are taken as auxiliary variables. The regression estimator based on the double-sampling approach is recommended here when the enumerator’s assessment of crop produce is taken as the main variable, while crop produce through the farmer recall method is taken as the auxiliary variable. The estimates based on both the simple- and the double-sampling approach exhibit a high degree of percentage CVs, which highlights the need to observe a larger sample for estimation purposes. The relatively poor performance of the double-sampling regression estimator compared to the linear estimator in maize crops is attributable to the small sample sizes used, coupled with high variability within the data.

The analysis on the crop yield presented above does not take into account the cost of data collection. However, the comparison between the double-sampling regression estimator and the linear estimator remains valid under the same cost or the same variance. In this study, the cost for a fixed value of variance of the estimator has been minimized. For the purposes of comparison, the CV of the estimator was fixed at 5 per cent, and the ratio between the cost incurred for collecting data on the main variable and the cost incurred for data collection on the auxiliary variable was considered 10 and 15. The cost function used here is the same as that used in Indonesia. The optimum values of the sample sizes were obtained using the same formulae as those used in Indonesia.

**Table 6.2.2.4. Estimates of crop yield with percentage CVs, using simple estimators and double-sampling regression estimators involving different variables**

Crop mixture name	Crop name	Crop yield by CCE		Crop yield by enumerator assessment of produce		Crop yield by CCE				Crop yield by Enumerators Assessment of produce			
						Farmer prediction of crop produce as auxiliary variable		Farmer recall of crop produce as auxiliary variable		Farmer prediction of crop produce as auxiliary variable		Farmer recall of crop produce as auxiliary variable	
		Est. (kg/h)	% CV	Est. (kg/h)	% CV	Est. (kg/h)	% CV	Est. (kg/ha)	% CV	Est. (kg/ha)	% CV	Est. (kg/ha)	% CV
Climbing beans	Climbing beans	338.2	60.3	565.1	24.2	425.3	40.7	513.2	33.7	623.6	18.6	682.6	16.9
Maize	Maize	2181	19.6	812.8	21.0	2,014.2	20.3	2,061.6	20.7	691.5	21.1	580.0	28.2
Climbing beans and maize	Maize	232.5	70.6	308.7	89.1	258.3	55.7	286.5	49.3	353.5	67.4	403.7	57.5

### 6.2.3 Jamaica

#### Crop area estimation

Figure 6.2.3.1 presents a scatter plot between parcel areas measured by GPS and by farmer inquiry. Unlike the case in Indonesia and in Rwanda, the relationship between the parcel areas measured by GPS and those determined through farmer inquiry is almost linear. The magnitude of  $R^2$  is equal to 0.741.

**Figure 6.2.3.1. Scatter plot indicating parcel area determined by GPS and through farmer inquiry**

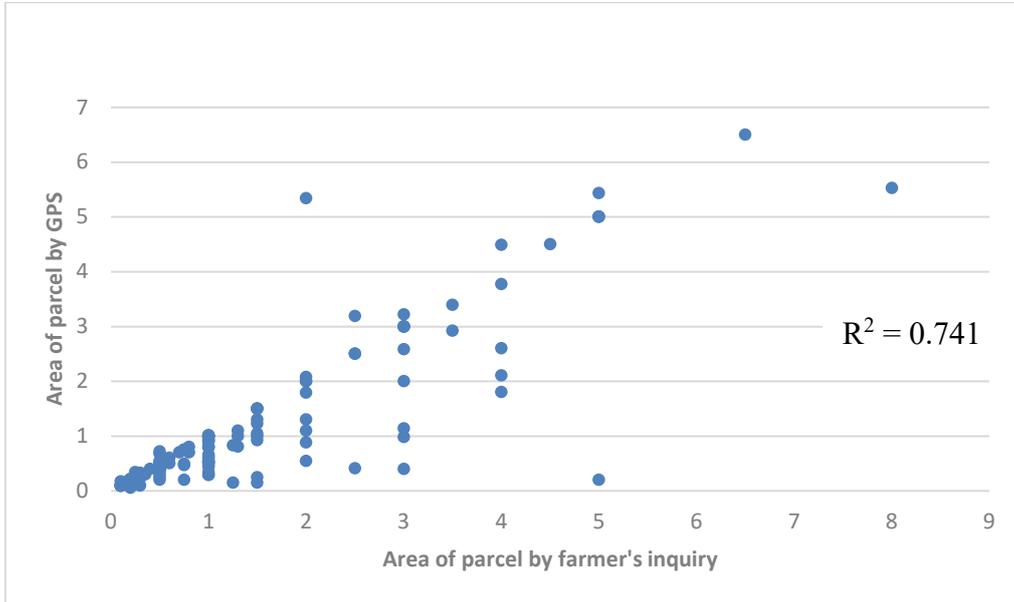


Figure 6.2.3.2 is a scatter plot between the sown area of the parcel, as measured by GPS and by means of farmer inquiry. It may be seen that a near-linear relationship exists between the sown areas of parcels as measured by GPS and by farmer inquiry. The magnitude of  $R^2$  is 0.798.

**Figure 6.2.3.2. Scatter plot indicating sown area of parcel, as measured by GPS and by means of farmer inquiry**

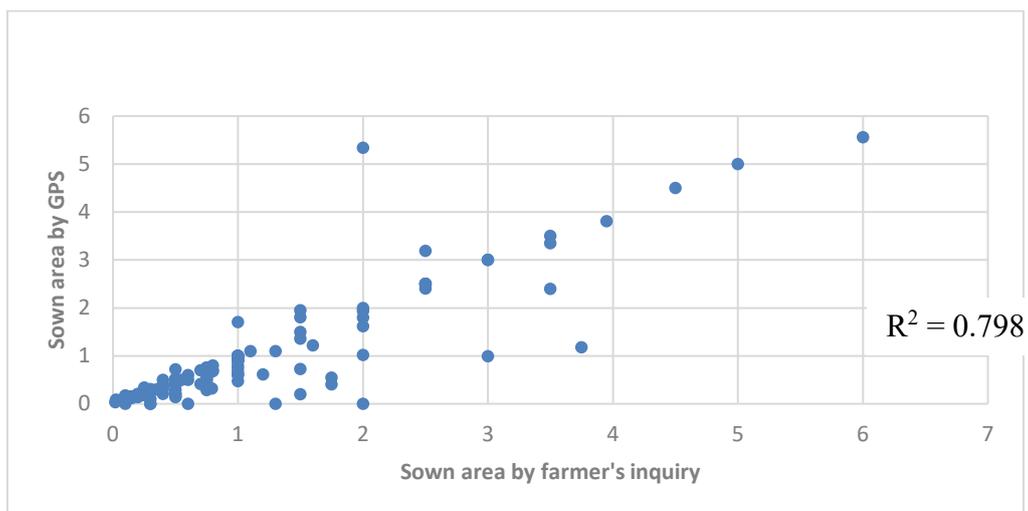


Table 6.2.3.1 provides the unweighted correlation matrix between total family members, active family members, parcel area by farmer inquiry and parcel area by GPS. There is a high correlation between parcel area by GPS and inquiry-based parcel area. The total number of family members and that of active family members exhibit a poor correlation with the parcel areas computed on the basis of GPS and of farmer inquiries.

**Table 6.2.3.1: Correlation matrix indicating the correlation between total family members, active family members, parcel area by inquiry and parcel area by GPS**

Variable	Total family member	Active family member	Area by inquiry	Area by GPS
Total family member	1.00	0.58	-0.03	-0.03
Active family member	0.58	1.00	0.03	-0.04
Area by inquiry	-0.03	0.03	1.00	0.86
Area by GPS	-0.03	-0.04	0.86	1.00

**Table 6.2.3.2: Category-wise average relative difference between parcel area measurement through GPS and through farmer inquiry, and the average relative difference between measurements of sown area through GPS and through farmer inquiry**

Area size class	Number of parcels	Average difference in parcel area and farmer inquiry	Average percentage relative difference in parcel area and farmer inquiry	Average difference in sown area and farmer inquiry	Average percentage relative difference in sown area and farmer inquiry
Overall	138	0.29	75.77	0.12	31.53
<= 1 ha	98	0.31	99.54	0.14	38.66
>1 ha & <= 2 ha	16	0.50	36.29	0.10	15.38
>2 ha & <= 3 ha	11	0.38	16.02	0.05	9.43
> 3 ha	13	-0.18	-4.30	0.07	17.87

A close look at Table 6.2.3.2 reveals that the average relative difference between the parcel areas and sown areas obtained through GPS and through farmer

inquiry decreases as the parcel area and the sown area increase. The differences between GPS and farmer inquiry are narrower for sown areas.

A close examination of Table 6.2.3.3 shows that for all three crops, the percentage CVs are lowest in the double-sampling regression estimator involving the GPS-based crop area as the study variable and the farmer inquiry-based crop area as the auxiliary variable. However, in all cases, the percentage CVs are higher. Thus, larger sample sizes are necessary to estimate crop area.

**Table 6.2.3.3. GPS-based linear estimates and double-sampling regression estimates using total number of family members, active family members and farmer inquiry-based crop area as auxiliary variables, along with percentage CVs**

Crop name	Crop-wise sample sizes of EDs	Crop Area Estimates							
		By GPS only		Using total number of family members as auxiliary variable		Using active family members as auxiliary variable		Using crop area by inquiry as auxiliary variable	
		$\hat{Y}_{GPS}$ (ha)	%CV	$\hat{Y}_{FM}$ (ha)	%CV	$\hat{Y}_{AFM}$ (ha)	%CV	$\hat{Y}_{INQ}$ (ha)	%CV
Irish potato	5	88.88	60.7788	69.99	76.2762	77.21	69.3449	300.06	14.6279
Sweet potato	5	125.69	73.4919	138.53	66.326	143.94	63.8553	259.88	35.3934
Yellow yam	18	3,160.16	47.9122	2,601.63	57.9594	2,898.1	51.9934	3,460.21	43.3933

### Crop yield estimation

In Jamaica, various methods of crop yield estimation were attempted in this study. In the absence of data on the Gold Standard, it is not possible to recommend one particular method. Furthermore, unlike the situation in Indonesia and Rwanda, only the linear estimator was considered for yield estimation. The optimum values of the sample sizes are also provided.

Figure 6.2.3.3 presents scatter plots between the crop yields obtained by farmer's prediction, farmer's recall, CCE, whole field harvest and sampling of harvest units.

**Figure 6.2.3.3. Scatter plots between the main variables and the auxiliary variables of methods of crop yield measurements, i.e. crop yield by farmer's prediction, farmer's eye estimates, CCE and sampling of harvest units.**

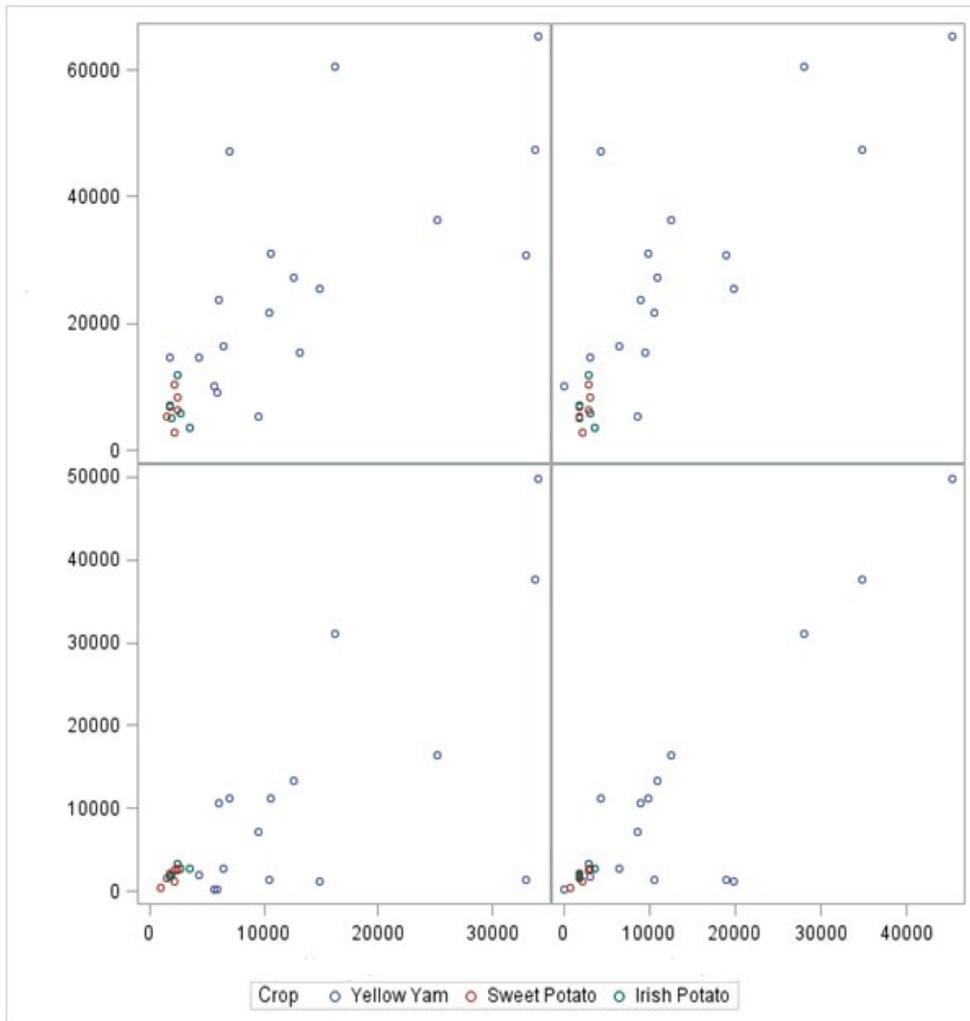


Table 6.2.3.4 provides linear estimates of crop yield based on CCEs, farmer assessments of crop produce, eye-estimated crop produce and sampling of harvest units, together with the respective percentage CVs for yellow yam and sweet potato. An estimator based on farmer assessments of crop produce is preferable to the other estimators, in terms of the criterion of the estimator's percentage CV. The CCE method tends to overestimate the crop yield. The regression estimator based on double sampling cannot be used as information on auxiliary variables, as farmer-predicted and farmer-recalled produce cannot be observed in larger samples. Higher percentage CVs are the result of smaller sample sizes.

**Table 6.2.3.4. Estimates of crop yield by CCE, farmer assessment (5A), eye estimates on day of harvesting, and sampling of harvest units, together with percentage CVs using a simple linear estimator.**

Crop name	CCE		Farmer assessment (5A)		Eye estimate		Sampling of harvest units	
	Est. (kg/ha)	% CV	Est.(kg/ha)	% CV	Est. (kg/ha)	% CV	Est. (kg/ha)	% CV
Yellow yam	27,502.04	27.27	13,716.28	27.73	16,247.37	31.22	11,010.00	47.50
Sweet potato* *	7,603.11	12.86	2,045.04	4.10	2,448.00	5.24	2,224.14	0.67

\*\* Estimates based on five parcels only

### 6.3. Results from World Bank methodological experiments in Africa on area measurement

Similar studies were carried out by the World Bank within the context of the LSMS conducted in Malawi (2010-11), Niger (2011), Uganda (2009-10) and the United Republic of Tanzania (2010-11) on larger samples, which reveal a moderate to high correlation between crop area determined by GPS and by farmer inquiry. The correlation coefficient between GPS and farmer self-reported plot areas are given in Table 6.3.1. The table shows that the correlations vary from moderate to high; this is different from the case of the present study, in which the magnitude of the correlation coefficient between GPS and farmer self-reported areas are very high.

**Table 6.3.1. Correlation coefficients between GPS and farmer self-reported plot areas in different countries.**

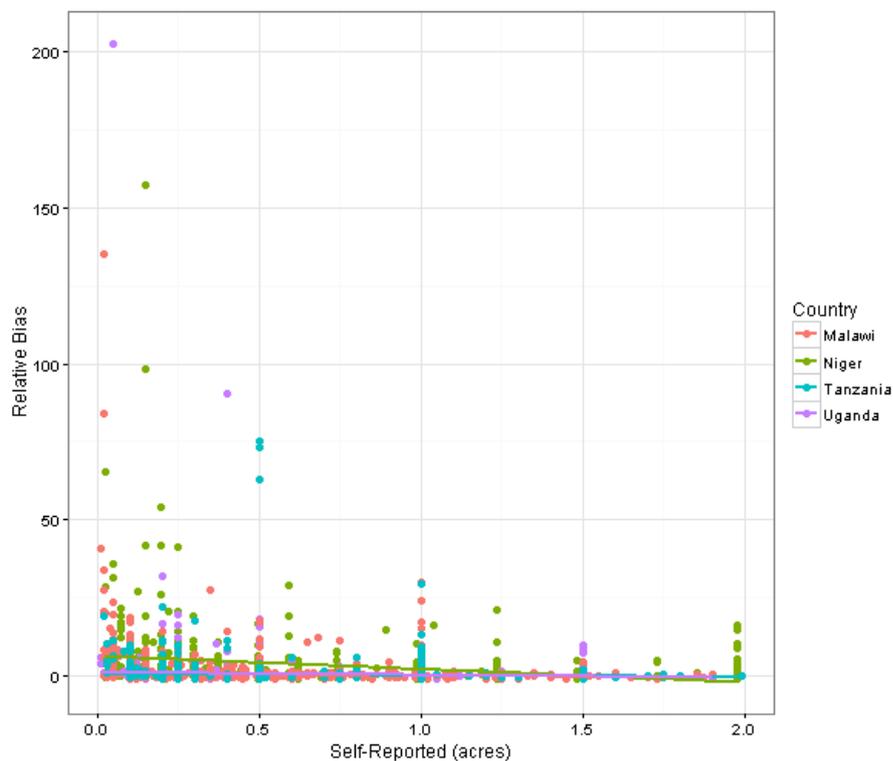
Country	Correlation coefficients between GPS and farmer self-reported plot areas
Niger	0.457
Uganda	0.771
Malawi	0.409
United Republic of Tanzania	0.797

Source: Carletto *et al.* (2015).

The scatter plot in Figure 6.3.1 illustrates the relationship between relative bias and farmer self-reported plot area (in acres) in four countries: Malawi (2010-11), Niger (2011), Uganda (2009-10) and United Republic of Tanzania (2010-11). The relative bias is the area measurement performed by GPS minus farmer self-

reporting divided by the farmer self-reported plot area. Only the areas smaller than 2.0 acres were considered, because over 80 plots in the scatter plot have an area smaller than 2.0 acres.

**Figure 6.3.1. Scatter plot indicating relationship between relative bias and self-reported plot area by farmers in Malawi (2010-11), Niger (2011), Uganda (2009-10) and United Republic of Tanzania (2010-11).**



One of the reasons for the discrepant results between the present study and the World Bank's study may consist in the sample sizes, which were moderate in the present study. Another possible reason is the difference in the locations chosen to conduct the two studies. Furthermore, the high literacy rate in Indonesia explains the high correlation between GPS and farmer self-reported plot areas (Carletto et al., 2016).

## Recommendations

The major recommendations formulated on the basis of the results of the field tests are the following:

- The use of list frames is recommended where up-to-date sampling frames of holders or farmers are not readily available and where villages/CB/EDs have distinct boundaries.
- The area frame approach is useful when the frame's component units are segments of land and village boundaries are uncertain. Area frames provide a complete coverage of land (GSARS, 2015a).
- The use of GPS methods are recommended where parcels are relatively larger in size and where high variability within the population prevents the use of the rope-and-compass method; this conclusion agrees with the findings of Keita *et al.* (2009).
- The method of sampling harvest units is useful when dealing with high-productivity crops. Given its simplicity, it is a useful method for estimating crop yield when sample sizes are very large (Fermont & Benson, 2011).
- The method relying upon enumerator self-reported parcel area is recommended in countries where the enumerators are knowledgeable and familiar with local conditions, and can thus objectively assess the dimensions of the parcel or of the sown area. De Groote and Traoré (2005) report that enumerators, in consultation with farmer, can provide accurate estimates of parcel size.
- Different methods of apportioning the crop mixture area into component crops exist. However, the use of objective methods (such as apportioning) based on measurements of plant density are expected to provide accurate estimates of the area under the component crops within the mixture.

- The method of enumerator assessment of crop produce is recommended for countries where enumerators are experienced and capable of objectively assessing the crop produce. Fermont and Benson (2011) observed that enumerators are often capable of estimating crop production or yield by visually assessing crop conditions, such as colour, plant vigour and plant density in the field.
- The method of crop yield estimation through farmer eye-estimated produce on the day of harvest is recommended for countries where farmers reside outside the villages or EDs and are involved in the cultivation of crops on a regular basis.
- It is recommended to adopt a combination of subjective and objective methods in countries where data collection on relevant auxiliary variables is relatively inexpensive, or where such auxiliary variables are readily available. Subjective and objective methods can be optimally combined through the use of double-sampling regression estimators (Sukhatme *et al.*, 1984).
- Use of the sample survey approach to estimate crop area and crop yield is most appropriate if the population size is large. The use of the stratification technique is expected to ensure a sample that is representative of the study population. Furthermore, the use of unequal probability sampling is expected to resolve the issues deriving from PSUs of unequal size (Cochran, 1977).
- The domain estimation approach enables framing of the mixture's component crop-wise estimators and their percentage standard errors. The approach is useful in mixed-crop-related scenarios (Sarndal *et al.*, 1992).

# Conclusion

On the basis of a gap analysis, this Technical Report attempts to address the problems relating to the estimation of crop area and crop yield in mixed and continuous cropping. Accordingly, an appropriate methodology has been developed to estimate crop area and crop yield in the contexts of mixed cropping and intercropping. To estimate the crop area of component crops in crop mixtures, the domain estimation approach has been proposed. The various crop mixtures are considered as domains. To estimate crop area and crop yield, the sample survey approach is proposed. Various measurement methods – both subjective and objective – to determine crop area and yield are explored, as well as their respective advantages and disadvantages, in light of the results of the field test.

The double-sampling approach has been used extensively to generate suitable estimators of crop area and crop yield. In estimation involving crop mixtures, an important issue is the apportioning of the crop area to the component crops. Standard procedures for apportioning crop areas in the mixture to the component crops, which take into account e.g. the seed rate, plant density or the number of rows, can be used for field crops. However, no such procedures are readily available for apportioning in mixtures involving annual and seasonal crops, annual and annual crops or – for that matter – mixtures of annual and perennial or perennial and perennial crops. Therefore, methods to apportion crop mixture areas to the component crops are provided to cover such cases.

Appropriate recommendations have been made on the application of the proposed sampling methodology, as well as on suitable methods to determine crop area and crop yield, bearing in mind the results of the field tests and the relevant literature available.

# References

- 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. George Washington University Capstone Report: Washington, D.C.
- Bradbury, D.** 1996b. Crops in Europe-Statistical systems for measuring area, production, and yield of non-cereal crops. Eurostat Publication: Luxembourg.
- Cochran, W.G.** 1977. Sampling Techniques. John Wiley & Sons: New York, USA.
- Carfagna, E.** 2004. List frames, area frames and administrative data, are they complementary or in competition?. Paper prepared for the Third International Conference on Agricultural Statistics, 2-4 November 2004. Cancún, Mexico.
- Carletto, G., Gourlay, S. & Winters, P.** 2015. From Guesstimates to GPStimates: Land Area Measurement and Implications for Agricultural Analysis. *Journal of African Economies*, 24(5): 593-628.
- Carletto, G., Gourlay, S., Murray, S. Zezza, A.** 2016a. Cheaper, Faster, and More Than Good Enough- Is GPS the New Gold Standard in Land Area Measurement? Policy Research Working Paper 7759. World Bank Publication: Washington, D.C.
- Carletto, G., Gourlay, S., Murray, S. & Zezza, A.** 2016b. Land Area Measurement in Household Surveys: A Guidebook. World Bank Publication: Washington D.C.
- Casley, D.J. & Kumar, K.** 1988. The collection, analysis and use of monitoring and evaluation data. Johns Hopkins University Press: Baltimore, Maryland, USA.
- 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.

- De Groot, H. & Traoré, O.** 2005. The cost of accuracy in crop area estimation. *Agricultural Systems*, 84(1): 21–38.
- Diskin, P.** 1997. *Agricultural productivity indicators measurement guide*. US Agency for International Development (USAID) Publication: Washington, D.C.
- Erenstein, O., Malik, R.K. & Singh, S.** 2007. Adoption and impact of zero-tillage in the rice-wheat zone of irrigated Haryana, India. Joint Publication of the International Centre for Maize and Wheat Improvement (CIMMYT) and the Rice-Wheat Consortium for the Indo-Gangetic Plains: New Delhi.
- Fermont, A.M. & Benson, T.** 2011. Estimating yield of food crops grown by smallholder farmers: A Review in the Uganda context. International Food Policy Research Institute (IFPRI) Publication: Washington, D.C.
- Gallego, F.J., Delincé, J. & Carfagna, E.** (1994). Two-Stage Area Frame Sampling on Square Segments for Farm Surveys. *Survey Methodology*, 20(2): 107-115.
- Gallego, F.J.** 1995. Sampling frames of square segments. Publication of the Office for Publications of the E.C.: Luxembourg.
- Gallego, F.J.** 2013. The use of a point sample as a master frame for agricultural statistics, Paper prepared for the Sixth International Conference on Agricultural Statistics, 23-25 October 2013. Rio de Janeiro, Instituto Brasileiro de Geografia e Estatística (IBGE).
- Gibson, J. & McKenzie, D.** 2007. Using the Global Positioning System (GPS) in Household Surveys for Better Economics and Better Policy. *World Bank Research Observer*, 22(2): 217-41.
- Global Strategy to improve Agricultural and Rural Statistics (GSARS).** 2015a. Technical Report on Linking Area and List Frames in Agricultural Surveys. Technical Report Series GO-08-2015. GSARS Technical Report: Rome. Available at <http://gsars.org/wp-content/uploads/2015/10/TR-on-Linking-Area-and-List-Frames-in-Agricultural-Surveys-191015.pdf>. Accessed on 10 February 2017.
- GSARS.** 2015b. Handbook on Master Sampling Frames for Agricultural Statistics: Frame Development, Sample Design and Estimation. GSARS Handbook: Rome. Available at

<http://gsars.org/wp-content/uploads/2016/02/MSF-010216-web.pdf>. Accessed on 10 February 2017.

**Hagblad, L.** 1998. Crop Cutting Versus Farmer Reports—Review of Swedish Findings. Statistik Rapport 1998: 2. Statistics Sweden: Örebro, Sweden.

**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. Michigan State University Publication: East Lansing, MI, USA.

**Just, R.** 1981. The productivity of mixed cropping practice: A study of survey data from Northern Nigeria. Parts I and II: West Africa Projects. World Bank Publication: Washington, D.C.

**Keita, N.** 2003. Root and tuber crops: Concepts and methods recommended by FAO and operational issues. Paper presented at the Expert Consultation on Root Crop Statistics, 3-6 December 2002. Harare.

**Keita, N. & Carfagna, E.** 2009 Use of modern geo-positioning devices in agricultural censuses and surveys. Bulletin of the International Statistical Institute. Paper prepared for the 57th Session of the International Statistical Institute, 16-22 August 2009. Durban, South Africa.

**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.

**Kelly, V., Hopkins, J., Reardon, T. & Crawford, E.** 1995. Improving the measurement and analysis of African agricultural productivity: Promoting complementarities between micro and macro data. International Development Paper No. 16. Michigan State University Publication: East Lansing, MI, USA.

**Kilic, T., Zezza, A., Carletto, G., & Savastano, S.** 2013. Missing(ness) in Action: Selectivity Bias in GPS-Based Land Area Measurements. World Bank Policy Research Working Paper 6490. World Bank Publication: Washington, D.C.

**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 for high

potential production systems of the National Resources Systems Programme Renewable Natural Resources Knowledge Strategy. Department for International Development Publication. HDRA: Coventry, UK.

**Republic of India.** Ministry of Statistics and Program Implementation (MoSPI). 2008. Manual for Crop Area and Production Statistics. MoSPI Publication: New Delhi.

[http://mospi.nic.in/sites/default/files/publication\\_reports/manual\\_area\\_crop\\_production\\_23july08\\_1.pdf](http://mospi.nic.in/sites/default/files/publication_reports/manual_area_crop_production_23july08_1.pdf)

**Murphy, J., Casley, D.J. & Curry, J.J.** 1991. Farmers' estimations as a source of production data. World Bank Technical Paper 132. World Bank Publication: Washington, D.C.

**Norman, D.W., Worman, F.D., Siebert, J.D. & Modiakgotla, E.** 1995. The Farming Systems Approach to Development and Appropriate Technology Generation. FAO Publication: Rome.

**Panse, V.G., Rajagopalan, M. & Pillai, S.S.** 1966. Estimation of crop yields for smaller areas. *Biometrics*, 22(2): 374-384.

**Poate, C.D.** 1988. A review of methods for measuring crop production from smallholder producers. *Experimental Agriculture*, 24(1):1-14.

**Rozelle, S.** 1991. Rural household data collection in developing countries: designing instruments and methods for collecting farm production data. Working Paper in Agricultural Economics 91-17. Cornell University Press: Ithaca, NY, USA.

**Sheaffer, C.C. & Moncada, K.M.** 2011. Introduction to Agronomy: Food, Crops, and Environment. Delmar Cengage Learning: Boston, MA, USA.

**Sud, U.C., Ahmad, T., Gupta, V.K., Chandra, H., Sahoo, P.M, Aditya, K. Singh, M. & Biswas, A.** 2015. Gap Analysis and Proposed Methodologies for Estimation of Crop Area and Crop Yield under Mixed and Continuous Cropping. Global Strategy Working Paper No. 4. Available at <http://gsars.org/wp-content/uploads/2015/12/WP-2-on-Improving-Methods-for-Estimation-of-Crop-Area-and-Crop-yield-under-Mixed-and-Continuous-Cropping-141215.pdf>. Accessed on 10 February 2'17

**Sukhatme, P.V. & Panse V.G.** 1951. Crop surveys in India II. Journal of the Indian Society of Agricultural Statistics, 3(1-2): 97-168.

**Särndal, C.E., Swensson, B. & Wretman, J.** 1992. Model-Assisted Survey Sampling. Springer-Verlag: New York, USA.

**Sukhatme, P. V., Sukhatme, B. V., Sukhatme, S. & Asok, C.** 1984. Sampling Theory of Surveys with Applications. Third edition. Iowa State University Press: Ames, IA, USA.

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

**Verma, V., Marchant, T. & Scott, C.** 1988. Evaluation of Crop-Cut Methods and Farmer Reports for Estimating Crop Production: Results of a Methodological Study in Five African Countries. Longacre Agricultural Development Centre Ltd: London.

# Glossary

**Area abandoned:** the part of the area intended or area planted that has been abandoned for raising crops or harvesting crops for different reasons

**Area damaged:** account of the loss due to the effect of unfavourable factors, such as floods, rain, winds, snowfall and insect attacks.

**Area frame:** a collection of well-defined land units that is used to draw survey samples.

**Area intended for planting (or sowing):** the area that the holders plan or intend to sow under various crops.

**Area tilled:** the part of arable land on which work has been done to make the land fit for raising crops at a given point of time.

**Area under cultivation:** the total area sown or planted; after the harvest, it excludes the ruined area resulting from, for example, natural disasters and calamities.

**Cluster:** the smallest unit into which the population can be divided is called an elementary unit of the population; a group of such elementary units is known as a cluster.

**Cluster sampling:** cluster sampling is a sampling procedure in which clusters are considered as sampling units and all elements of the selected clusters are enumerated.

**Continuous cropping:** continuous crops, also known as successive crops, sequential crops or catch crops, are crops that are sown and harvested from a piece of land previously occupied by another crop, or even by the same crop, during the same agricultural year. Continuous cropping can take different forms:

**Continuing planting/harvesting:** repeated planting and harvesting of crops at particular intervals of time in an agricultural year.

**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).

Some other forms of continuous cropping are found by:

**Enlarging gradually (at given intervals of time):** the area of land planted to one or several crops.

**Replanting the same crop** on the same land after it has been damaged (totally or partially) by natural or other causes.

**Crop area:** the horizontal projection of a particular extent of the Earth's surface

**Crop yield:** the concept of crop yield is generally used to represent the average amount of produce obtained per unit of the crop area. In case of tree crops, the concept of yield covers the average amount of produce per tree.

**Domain:** in many surveys, estimates may be required for each of the classes or subpopulations into which the population is subdivided. The term "domain" refers to these subpopulations.

**Estimator:** an estimator is a statistic obtained by a specified procedure for estimating a population parameter. The estimator is a random variable and its value differs from sample to sample.

**Harvest year:** the calendar year in which the harvest begins.

**Harvested area:** the part of the sown or planted area that is harvested.

**Intercropping:** the practice of growing more than one crop in the same land area, in rows of definite proportion and pattern.

**Mixed intercropping:** growing two or more crops simultaneously on the same piece of land with no distinct row arrangement.

**Row intercropping:** growing two or more crops simultaneously where one or more crops are planted in rows.

**Strip intercropping:** growing two or more crops simultaneously in different strips that are wide enough to carry out independent cultivation, but narrow enough for the crops to interact agronomically.

**Relay intercropping:** a technique in which different crops are planted during different time periods in the same field and both (or all) crops are grown simultaneously at least part of the time.

**List frame:** a list of units that can be sampled.

**Main area of a given parcel:** the area where the parcel has been used only once during a given crop year.

**Mixed cropping:** mixed crops refer to two or more different temporary and permanent crops grown simultaneously in the same field or plot.

**Monocropping:** the practice of growing only one crop on a piece of land year after year.

**Non-sampling error:** errors other than sampling errors such as those arising through non-response, incompleteness and inaccuracy of responses or measurements; these are likely to be more common and significant in a complete enumeration survey than in a sample survey.

**Percentage Standard Error:** the percentage standard error of the estimator  $t$  is defined as  $\%SE(t) = \frac{SE(t)}{t} \times 100$ .

**Population:** the collection of all units of a specified type in a given region at a particular point or period of time.

**Population parameter:** any function of the values of all population units (or of all observations constituting a population).

**Primary crops:** crops that come directly from the land without having undergone any real processing apart from cleaning. These can be further divided into temporary and permanent crops:

**Permanent crops:** crops that are sown or planted once and need not be replanted after each annual harvest.

**Temporary crops:** crops that are sown and harvested during the same agricultural year, sometimes more than once.

**Probability Sampling:** method of sampling in which the units in the sample are selected by some probability mechanism.

**Production area:** in connection with permanent crops, the area that can potentially be harvested in the reference harvest year.

**Random sample:** a sample the selection of which is governed by ascertainable laws of chance.

**Sample:** a fraction of the population or a proper subset of the population.

**Sampling error:** the error that arises due to inferences drawn on the population parameter on the basis of observations on a part (or sample) of the population.

**Sampling frame:** a list of all the sampling units belonging to the population to be studied with their identification particulars or a map showing the boundaries of the sampling units.

**Sampling unit:** elementary units or group of such units which besides being clearly defined, identifiable and observable, are convenient for the purpose of sampling.

**Sampling variance:** a measure of sampling error; the sampling variance of an estimator  $t$  is a measure of the difference between the estimated values and its expected value.

**Simple Random Sampling without Replacement (SRSWOR):** the simplest method of probability sampling. In this method, the units are drawn one by one, assigning an equal probability of selection to each unit at each draw; the unit selected on any particular draw is not returned to the population before selecting a unit at the next draw.

**Sole crop:** a crop grown in pure stand.

**Sown or planted area:** the area that corresponds to the total sown area for producing a specific crop during a given year.

**Standard error:** the positive square root of variance is termed as standard error of the estimator  $t$ , i.e.  $SE(t) = \sqrt{Var(t)}$ .

**Statistic:** any function of values of sample observations that is free from unknown population parameters.

**Strata:** stratification consists of dividing the population into homogeneous subgroups, subpopulations or subsets.

**Stratified sampling:** a commonly used technique of sampling where the population is divided into non-overlapping homogeneous subgroups or subpopulations called strata, which together comprise the entire population; an independent sample is then drawn from each stratum.

**Two-phase sampling:** a sampling technique that involves sampling in two phases (occasions). This technique is also referred to as double sampling.

**Two-stage sampling:** type of sampling that consists of first selecting clusters and then selecting a specified number of elements from each cluster chosen.

**Utilized agricultural area:** total area taken up by arable land, permanent grassland, permanent crops and kitchen gardens used by the holdings, regardless of type of tenure or use as common land.