IMPROVING PADDY RICE STATISTICS USING AREA SAMPLING FRAME TECHNIQUE

Anna Christine Durante, Pamela Lapitan, David Megill, and Lakshman Nagraj Rao

NO. 565

November 2018

ADB ECONOMICS WORKING PAPER SERIES

ASIAN DEVELOPMENT BANK

ADB

ADB Economics Working Paper Series

Improving Paddy Rice Statistics Using Area Sampling Frame Technique

Anna Christine Durante, Pamela Lapitan, David Megill, and Lakshman Nagraj Rao

No. 565 | November 2018

Anna Christine Durante (adurante.consultant@adb.org) is a consultant, Pamela Lapitan (plapitan@adb.org) is an Associate Economics and Statistics Officer, David Megill (davidmegill@yahoo.com) is a consultant, and Lakshman Nagraj Rao (NagrajRao@adb.org) is a Statistician, all from the Economic Research and Regional Cooperation Department of the Asian Development Bank.

This study was carried out under Regional Technical Assistance (R-CDTA) 8369: Innovative Data Collection Methods for Agricultural and Rural Statistics with the support of the Japan Fund for Poverty Reduction (JFPR). The authors benefited from the comments of Rana Hasan, Jesus Felipe, Kaushal Joshi, Valerie Mercer-Blackman, Takashi Yamano, Mahinthan Joseph Mariasingham, David Anthony Raitzer, Elisabetta Gentile, Kathleen Farrin, and other staff who participated in the ERCD seminar workshop. The authors are also grateful to the Ministry of Agriculture and Forests, Lao People's Democratic Republic; the Philippine Statistics Authority; Office of Agricultural Economics, Thailand; Ministry of Agriculture and Rural Development, Viet Nam; Remote Sensing Technology Center of Japan; and the Japan Aerospace Exploration Agency for providing the needed data in this study. Lea Rotairo, Rea Jean Tabaco, and Chrysalyn Gocatek provided excellent research assistance for this study.

ASIAN DEVELOPMENT BANK

ADB

Creative Commons Attribution 3.0 IGO license (CC BY 3.0 IGO)

© 2018 Asian Development Bank
6 ADB Avenue, Mandaluyong City, 1550 Metro Manila, Philippines
Tel +63 2 632 4444; Fax +63 2 636 2444
www.adb.org

Some rights reserved. Published in 2018.

ISSN 2313-6537 (print), 2313-6545 (electronic)
Publication Stock No. WPS189643-2
DOI: http://dx.doi.org/10.22617/WPS189643-2

The views expressed in this publication are those of the authors and do not necessarily reflect the views and policies of the Asian Development Bank (ADB) or its Board of Governors or the governments they represent.

ADB does not guarantee the accuracy of the data included in this publication and accepts no responsibility for any consequence of their use. The mention of specific companies or products of manufacturers does not imply that they are endorsed or recommended by ADB in preference to others of a similar nature that are not mentioned.

By making any designation of or reference to a particular territory or geographic area, or by using the term "country" in this document, ADB does not intend to make any judgments as to the legal or other status of any territory or area.

This work is available under the Creative Commons Attribution 3.0 IGO license (CC BY 3.0 IGO) https://creativecommons.org/licenses/by/3.0/igo/. By using the content of this publication, you agree to be bound by the terms of this license. For attribution, translations, adaptations, and permissions, please read the provisions and terms of use at https://www.adb.org/terms-use#openaccess.

This CC license does not apply to non-ADB copyright materials in this publication. If the material is attributed to another source, please contact the copyright owner or publisher of that source for permission to reproduce it. ADB cannot be held liable for any claims that arise as a result of your use of the material.

Please contact pubsmarketing@adb.org if you have questions or comments with respect to content, or if you wish to obtain copyright permission for your intended use that does not fall within these terms, or for permission to use the ADB logo.

Notes:
In this publication, "$" refers to United States dollars.
ADB recognizes "Vietnam" as Viet Nam.
Corrigenda to ADB publications may be found at http://www.adb.org/publications/corrigenda.

CONTENTS

TABLES AND FIGURES iv

ABSTRACT v

I. INTRODUCTION 1

II. STUDY AREAS 4

III. DATA DESCRIPTION 4
A. Sample Design 4
B. Plot Level Paddy Rice Area Estimation 10

IV. WEIGHTING, SAMPLING ERROR, AND DESIGN EFFECTS 11
A. Weighting Procedure 11
B. Calculation of Sampling Errors and Design Effects 13

V. RESULTS 14
A. Comparative Analysis of Direct Estimates of Total Area Planted in Rice 14
B. Yield Estimates Derived from Crop-Cutting Exercise 18
C. Direct Estimates of Total Production of Rice Paddy 20
D. Obtaining Mesh Level Estimates of Total Area Planted in Rice from Modified Map 20
E. Comparison of Area Estimates from Different Data Sources 23
F. Improving the Precision of the Estimates of Total Rice Paddy Production through Ratio Estimation 26

VI. CONCLUSION 28

REFERENCES 31

TABLES AND FIGURES

TABLES

1	Distribution of Meshes in the Sampling Frame for Each Pilot Province	6
2	Number of Sample Meshes Surveyed and Number of Meshes with Rice by Stratum	6
3	Standard Error, Coefficient of Variation, and Design Effects of Estimates of Total Area Planted in Rice Paddy Based on Area of Sample Plots from Unmodified Track Data	15
4	Standard Error, Coefficient of Variation, and Design Effects of Estimates of Total Area Planted in Rice Paddy Based on Area of Sample Plots from Modified Track Data	15
5	Estimates of Statistical Parameters for the Unweighted Paddy Area Data by Source	17
6	Estimate of Mean Yield of Rice Paddy per Subplot	18
7	Estimate of Mean Yield	19
8	Direct Estimates of Total Production of Rice Paddy	20
9	Estimate of Total Area Planted in Rice Paddy Based on Independent Measure of Total Area Planted with Rice in Each Sample Mesh, Using Digitized Google Earth Images	25
10	Ratio Estimate of Total Rice Paddy Production in the Pilot Areas	27

FIGURES

1	Sample 200 m x 200 m Mesh	5
2	An Example of a 200 m x 200 m Mesh on Printed Map with Rice Plots Identified and Serially Numbered within the Mesh	7
3	Numbering of Corners of Randomly Selected Plots	8
4	Measuring of Adjacent Sides of the Selected Corner	8
5	Plotting of Randomly Selected Distance from the Longer Arm	9
6	Plotting of Randomly Selected Distance from the Shorter Arm	9
7	Identifying the Location of the 2.5 m x 2.5 m Subplot	10
8	Unmodified Global Positioning System Track Data on Google Earth	11
9	Modified Track Data on Google Earth	11
10	Paddy Area Estimates in Savannakhet, Ang Thong, and Thai Binh Based on Data from Unmodified and Modified Track Files	17
11	Average Yield per Pilot Area	19
12	Total Production per Pilot Province by Stratum	21
13	Mesh Information on a Printed Map	22
14	Estimates of Total Planted Area in Savannakhet from Different Data Sources	23
15	Estimates of Total Planted Area in Ang Thong from Different Data Sources	24
16	Estimates of Total Planted Area in Thai Binh from Different Data Sources	24

ABSTRACT

Traditional sampling strategies for paddy rice statistics rely on outdated list frames, incomplete holding information, or administrative data that are prone to numerous biases. The objective of this study is to test the utility of an area frame developed using remote sensing data in three pilot provinces—Savannakhet (Lao People's Democratic Republic), Ang Thong (Thailand), and Thai Binh (Viet Nam). Direct estimates of total paddy rice area and production are calculated from area frame using two methods––one involving measurement of plot size using a Global Positioning System instrument and the other utilizing a digitized map of farmer-identified plot boundaries on a high-resolution Google Earth image. A third method involving the calculation of ratio estimates using independent mesh-level measures is compared with the first two methods involving direct estimates, and with the estimates generated from administrative data from the countries. Our study finds that ratio estimation significantly improves the level of precision of paddy rice statistics. Substantial deviations are also observed between official statistics and the statistics generated through direct estimation.

Keywords: agriculture, sampling methods

JEL codes: C83, O13, Q19

I. INTRODUCTION

Timely and reliable agricultural statistics are critical for monitoring government agricultural development plans and mitigating the effects of extreme weather and climate change. They are also useful in gaining a better understanding of people's well-being through timely and effective policy interventions. The preparation of national accounts, evaluation of agricultural interventions, and the development of early warning systems to address climatic and nonclimatic vulnerabilities in the agriculture sector, rely on high-quality and disaggregated agricultural data. In the absence of good quality data, inefficient allocation of resources is likely which would lead to a failure in resolving critical development problems (Kelly et al. 1995).

The compilation of official agricultural statistics relies on data collected using administrative records or probability-based field surveys. In the case of administrative data, the starting point in most cases is a government agricultural officer who determines crop area and production in his or her assigned locality by observing harvests and interviewing experts such as village heads, farmers, and traders. These estimates are reported to the next level of bureaucracy until the summary statistics reach the national government (ADB 2016). The advantage of this method lies in its lower implementation cost, but estimates derived are likely to be biased and prone to large measurement errors. Data collection officers and others involved in the process may have vested interests to either support their claims of accomplishment that influence the estimation process upward or showcase a downward trend in expectations of subsidies or other government amenities (Carfagna and Carfagna 2010).

When objectively designed and conducted, household and/or agricultural surveys can provide better estimates. However, measurement errors may still arise because methodological studies suggest that during interviews, farmers, for a multitude of reasons, may inadvertently provide inaccurate crop areas and production estimates. Firstly, the accuracy of subjective estimates may be a function of respondent characteristics. A study by Carletto, Savastano, and Zezza (2013) found that more educated farmers provided better estimates of area while absentee landlords and respondents for whom farming is a secondary activity were less informed about plot characteristics. The quality of data collected through farmer self-reporting may also be significantly affected by the predisposition of respondents to round off areas and their misunderstanding of the existence of property rights and plot condition such as slope and crop type (Carletto, Gourlay, and Winters 2015; De Groote and Traoré 2005). Finally, lack of cadastral information on land use, intercropping, postharvest losses, nonuniformity of plots can all pose challenges in estimating key crop-related statistics (Carfagna and Carfagna 2010). Few studies have compared estimates from administrative sources and surveys and have found significantly different results (ADB 2016; Beegle, Carletto, and Himelein 2012; Sandefur and Glassman 2014; Deininger et al. 2011).

Both agricultural and population-based census frames are commonly used in developing countries as a basis for designing multistage sample agricultural and household surveys (Grosh and Munoz 1996). In the first sampling stage, primary sampling units are selected from census enumeration areas. In the second stage, agricultural households are selected from a housing frame which is generated through field listing activities. However, in some countries, a complete frame is not available if the reference is a census with low coverage, or the existing lists of sampling units change rapidly rendering the list frame out-of-date (Griffin 2014). Field listing activities may not be accurate if households are systematically overreporting or underreporting agricultural holdings. Semi-nomadic households engaged in agriculture that are temporarily absent or fully nomadic households without

fixed dwellings are also undercovered by this approach leading to substantial biases in agricultural statistics (Himelein, Eckman, and Murray 2014).

An alternative to the list frame approach is the area frame approach. In the area frame approach, the final stage sampling units are land areas and the selection probabilities are proportional to their area measures. The geographic scope of the region of interest is identified and divided into nonoverlapping units (Biemer 2010). The units are defined by a geometric grid, usually points, squares, circles.[1] A multistage stratified approach can then be implemented based on an area frame to select a sample of grids within each stratum of land cover and/or land use, depending on the survey objective (Faulkenberry and Garoui 1991). The United States Department of Agriculture has been utilizing the area frame approach for estimating agricultural production and livestock statistics since the 1930s (Davies 2009). Several countries in Europe have also used segments with regular geometric shapes for estimating crop area statistics through the MARS and LUCAS projects (Fuentes and Gallego 1994, Gallego 2007). Much of this work has been facilitated through advancement in remote sensing and geographic information system (GIS) techniques and has led to savings in cost and time (Khan et al. 2010, Atzberger 2013, Cotter et al. 2010, Strand 2013, Boryan et al. 2017, Carfagna and Gallego 2005).

Despite the mainstreaming of the area frame approach using geometric grids in many advanced economies, most developing countries in Asia and the Pacific continue to rely on traditional surveys or administrative data methods to generate crop statistics, with the exception of a few pilot studies (Singh et al. 1992, Singh et al. 2002). Two major reasons can be cited for area frames not being used: (i) the lack of recent and accurate cadastral maps, and (ii) personnel with limited skills in remote sensing and GIS in national statistics offices in the region. To fill in the gap in the existing literature, this study utilizes an area frame approach through the innovative combination of satellite data, GIS methods, and crop cutting to estimate paddy rice area, yield, and production for the 2015 rainy season in selected provinces of three pilot countries—the Lao People's Democratic Republic (Lao PDR), Thailand, and Viet Nam—and compares them to estimates obtained from existing administrative data sources.[2]

The pilot provinces were stratified into rice-growing areas using satellite data and GIS methods and within each stratum, square meshes were randomly selected to identify plots eligible for crop cutting. Crop cutting was then implemented in randomly selected subplots in each sample plot to obtain unbiased rice yield estimates. This allowed for the calculation of both direct and indirect estimates of total paddy rice area. A third method to estimate total paddy rice area was also tested, which involves the calculation of ratio estimates through the measurement of total area planted in paddy rice based on independent mesh-level measures using the Agricultural Land Information System (ALIS) methodology. The level of precision obtained from this method is finally compared with the two direct estimates and administrative data provided by the counterpart government agencies in each country.

---

[1] Historically, paper-based cadastral maps of nongeometric and nonoverlapping plots have been used as an alternative to geometric grids in countries such as India and Bangladesh for crop-cutting surveys. Such maps would serve as the gold standard for implementing agricultural surveys, as long as these are updated routinely and accurately, with the use of GIS techniques. In most countries in Asia and the Pacific, such cadastral maps are unavailable, outdated, or of low quality.

[2] The pilot provinces include: Savannakhet (Lao PDR), Ang Thong (Thailand), and Thai Binh (Viet Nam). Although the project was also implemented in Nueva Ecija (Philippines), the results are not presented here due to the occurrence of Typhoon Lando (international name "Koppu"), which did not allow the completion of field activities.

Our main findings from this study are as follows: First, the direct estimates of the total rice paddy area and production from the sample plots have relatively high coefficient of variations (CVs) and wide confidence intervals. In comparison, the level of precision of total rice paddy production was significantly improved through ratio estimation, where the ratio estimate of the rice yield per area was applied to a separate, more accurate estimate of the area planted in rice. The main reason for the relatively high CVs for the direct estimates of the total area and production of rice paddy is the variability in the size of the sample plots, which are selected within each sample mesh with equal probability.[3] This independent measure of total area planted in rice paddy at the sample mesh level reduces the sampling error that results from the variability in sample plot sizes and therefore provides a more precise estimate of total area planted in rice.

Second, we find significant deviations between official statistics, which are collected from administrative recording systems using farmer recall techniques in all three countries, and the estimates obtained from our study for the same cropping season.[4] For example, total area planted using the ratio estimate for Thai Binh was about 9.5% lower than official estimate obtained from the General Statistics Office of Viet Nam (GSO 2015). The ratio estimate for total paddy rice production in Savannakhet is about 19.7% lower than official estimates obtained from the CIS (CIS 2014). The yield estimates for Ang Thong and Thai Binh from crop cutting are close to official estimates but are about half the official estimate for Savannakhet. Although it is difficult to pinpoint the exact reasons for these deviations as the microdata from the administrative records were not made available to us, literature suggests that the presence of nonsampling errors, subjective intervention, and political leadership at the local government levels involving subsequent revisions in the administrative data could all be plausible explanations (ADB 2016).

Finally, although sufficient GIS data is freely available to undergo the stratification process, there were a few instances where the satellite data indicated rice, but no rice was found during the field validation. The opposite scenario was observed in a few other sample meshes. There are two possible explanations for this: (i) the power of discrimination in the satellite imagery and stratification might not be sufficient or (ii) field teams might not have accurately reported the status of all meshes, thereby systematically excluding some rice-growing meshes from the survey.

The rest of the paper is organized as follows. Section II describes the study areas, while section III presents the numerous datasets used in this study and fieldwork implementation. In section IV, we present the weighting technique implemented in this study, while section V showcases the key results. The concluding section provides recommendations for scale-up of the methodology in the pilot countries.

---

[3] The reason for implementing equal probabilities of selection for all plots despite varying size was the unavailability of reliable plot size information during the fieldwork. Not all plot owners were available at the site to get farmer estimates (which by itself presents measurement error concerns). Village officials designated to assist with the project did not maintain reliable records in all countries. Measuring the size of all plots within each sample mesh would escalate fieldwork implementation costs.

[4] In the Lao PDR and Viet Nam, the administrative reporting system relies on village/commune officials collecting data from farmers through farmer recall methods which are summarized and reported at every stage of the hierarchy (district, province, and national). In Thailand, annual sample surveys are implemented to obtain paddy rice statistics.

II. STUDY AREAS

The first study area, Savannakhet, is the Lao PDR's largest province with a total land area of 2,177,400 hectares (ha). The province exhibits a tropical wet and dry climate with very warm weather throughout the year. A large portion of its population is engaged in subsistence agricultural activities. Savannakhet was selected as part of this study in consultation with the government counterparts due to its substantial extent of contiguous paddy rice area. It is one of the most important rice-producing provinces in the country, accounting for nearly 23% of the national wet season paddy planted area in 2014 (MAF 2015).

Second, Ang Thong Province, which is in Thailand's central region, has a land area of 96,840 ha. It is mostly flat and consists primarily of agricultural land. Ang Thong is a key paddy rice production area in the central region, with approximately 58% of land in the province dedicated to paddy rice farming. Paddy rice is grown twice per year: once during the rainy season from May to July and another during the dry season from December to February (FAO 2012). Ang Thong was also selected as part of this study in consultation with the government counterparts due to its substantial extent of contiguous paddy rice area.

Third, Thai Binh is a coastal eastern province in Northern Viet Nam's Red River Delta region. The topography in Thai Binh is mostly flat with an average height of 1–2 meters (m) above sea level. The province has access to three rivers—Red, Luoc, and Hoa, making the soil very fertile. Complemented with adequate rainfall, the province provides an ideal setting for growing paddy rice. In fact, Thai Binh is acknowledged to grow the best strain of rice in Northern Viet Nam and as per official statistics, contributes to 2.7% of total rice production in Viet Nam (FAO 2002).

III. DATA DESCRIPTION

A. Sample Design

An area frame was used for this study and constructed based on the expected likelihood of finding paddy rice area in each grid square mesh. Two sources of rice maps were utilized to implement the stratification process: (i) rice extent maps using 2015 MODIS data produced by the International Rice Research Institute (IRRI)[5] and (ii) land use maps from 2009 produced by the European Space Agency (ESA) under its GLOBCOVER initiative.[6] These two sources allow for identification of land most recently used for growing rice alongside providing information on those areas which are repeatedly used for rice cultivation. The primary sampling unit in this study is a 200 m x 200 m square "mesh" which is spatially defined on a digitized satellite image map (Figure 1).[7]

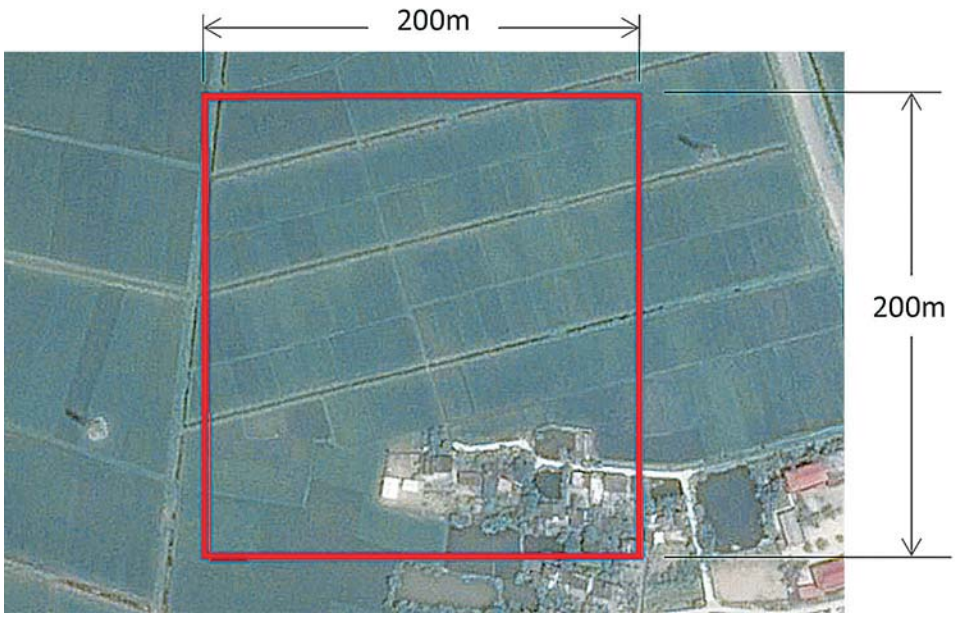
---

[5] IRRI has been developing remote sensing-based maps of rice systems in Asia as part of its contribution to various projects that need good baseline data on rice (http://irri.org/our-work/research/policy-and-markets/mapping/remote-sensing-derived-rice-maps-and-related-publications).

[6] GlobCover is an ESA initiative which began in 2005 in partnership with the Joint Research Center (JRC of the European Commission), United Nations Environment Programme, Food and Agriculture Organization of the United Nations, and other institutions. The aim of the project was to develop a service capable of delivering global composites and land cover maps using as input observations from a sensor onboard the Environmental Satellite (ENVISAT) mission (http://due.esrin.esa.int/page_globcover.php).

[7] The choice of 200 m x 200 m mesh is based on pixel size of the satellite images used in the study.

Figure 1: Sample 200 m x 200 m Mesh



m = meter.
Source: Google Earth. 13 April 2015. Thai Binh, Viet Nam. 20°29'52.15" N, 106°17'55.86" E, eye alt 709m. http://www.earth.google.com (accessed 13 October 2016).

The stratification in this study was conducted prior to the selection of meshes to improve statistical efficiency and lower fieldwork costs. The first stratum consisted of meshes that both IRRI and ESA maps identified as paddy rice area, considered to be the most likely to contain paddy rice. This stratum is henceforth referred to as the *IRRI+GlobCover* stratum. The second stratum consisted of meshes that were only identified as rice by the IRRI area map but not by the ESA map, henceforth referred to as the *IRRI* stratum. This was considered a medium probability stratum since the resolution of the IRRI map obtained from MODIS is better than the ESA map obtained from ENVISAT, and IRRI's classification is more recent than ESA.[8] The third stratum is the low probability stratum, identified as rice by ESA's map but not by IRRI's map, henceforth referred to as the *GlobCover* stratum. The final stratum consists of all remaining areas where presumably no rice is grown as indicated by both IRRI and ESA maps, henceforth referred to as the *Other* stratum. Therefore, within each stratum, the entire area was conceptually divided systematically into 200 m by 200 m meshes using GIS techniques. In this case, the number of meshes in each stratum would be equal to the total area of the stratum divided by 40,000 square meters (m^2).

In the first sampling stage, a stratified random sample of 120 meshes was selected for each pilot province.[9] A random sample of reserve meshes that could be used for possible replacement was also selected in each stratum. A sample mesh would only be replaced in extreme cases such as problems with security or accessibility. Also, the number of selected meshes was higher in the stratum

[8] The spatial resolution of MODIS is 250 m while the spatial resolution of ENVISAT is 300 m. Also, the IRRI map is more recent, and uses satellite data from 2015 while the GlobCover map is constructed using data from 2009.

[9] The total number of meshes was based on the expected number of rice plots to be found and interviewed in each stratum using data from pretests and the available budget for the pilot project.

where the expected likelihood of finding rice-growing plots is highest (Stratum 1), and lower in areas with low (Stratum 3) or no likelihood (Stratum 4) of finding rice-growing plots. The distribution of the total number of meshes in the frame by stratum and sample replacement meshes selected for Savannakhet, Ang Thong, and Thai Binh is shown in Table 1.

Table 1: Distribution of Meshes in the Sampling Frame for Each Pilot Province

Stratum	Sample Meshes Selected	Replacement Meshes Selected	Number of Meshes in Frame		
			Savannakhet	Ang Thong	Thai Binh
IRRI+GlobCover	80	5	80,839	22,105	36,376
IRRI	20	10	4,650	280	589
GlobCover	15	10	154,227	2,777	4,846
Others	5	5	322,391	34	1,815
Total	120	30	562,107	25,196	43,626

Notes: IRRI refers to meshes identified as paddy area by the International Rice Research Institute (IRRI) rice maps. GlobCover refers to meshes identified as paddy area by the European Space Agency (ESA) GlobCover land use maps. IRRI+GlobCover refers to meshes identified as paddy area by both IRRI and ESA area maps.
Source: Authors' estimates.

A ground-truthing field operation was conducted to verify whether rice was planted in any plots within the boundaries of each sample mesh. Only sample meshes with rice were enumerated for eligibility to be selected for crop cutting. The final distribution of the sample meshes with rice planted in Savannakhet, Ang Thong, and Thai Binh is shown in Table 2.

Table 2: Number of Sample Meshes Surveyed and Number of Meshes with Rice by Stratum

Stratum	Sample Meshes Surveyed			Sample Meshes with Rice		
	Savannakhet	Ang Thong	Thai Binh	Savannakhet	Ang Thong	Thai Binh
IRRI+GlobCover	80	79	79	58	50	63
IRRI	18	20	20	8	10	2
GlobCover	16	15	15	10	6	7
Others	5	5	5	2	0	0
Total	119	119	119	78	66	72

Notes: IRRI refers to meshes identified as paddy area by the International Rice Research Institute (IRRI) rice maps. GlobCover refers to meshes identified as paddy area by the European Space Agency (ESA) GlobCover land use maps. IRRI+GlobCover refers to meshes identified as paddy area by both IRRI and ESA area maps. For Savannakhet, several samples were inaccessible due to flooding at the time of the first field validation. These sample meshes were replaced accordingly. For Thai Binh, one sample mesh under the IRRI+GlobCover stratum was assigned to two different teams of enumerators, hence missing one sample mesh under the same stratum.
Source: Authors' estimates based on field validation of Center for Agricultural Statistics, Lao PDR; Office of Agricultural Economics, Thailand; and Center for Informatics and Statistics, Viet Nam.

For the second sampling stage, a listing of all rice plots identified with at least part of their area within the boundaries of each sample mesh was conducted. All plots where rice would be harvested during the rainy season of 2015 were eligible for selection at the second sampling stage. Plot boundaries were defined based on the definition adopted by the Living Standards Measurement Study Group of the World Bank, where a "plot" is "a continuous piece of land on which a unique crop or a mixture of crops is grown, under a uniform, consistent crop management system, not split by a path of more than 1 m in width, and with boundaries defined in accordance with the crops grown and the operator" (Kilic, Yacoubou Djima, and Carletto 2017).

A printed map of each of the 200 m x 200 m sample meshes was used to identify the number of rice plots within each mesh. Landmarks on the printed map were matched with what is observed on the field. Boundaries of the mesh were verified using a Global Positioning System (GPS) application installed on the handheld device used for fieldwork, which showed the field staff's current position in relation to the mesh. The plot boundaries and the respective owners were identified with the help of the village heads. After the boundaries of all the plots were identified and delineated on the printed map, each plot was numbered in a geographically serial and serpentine manner (Figure 2). A listing form was used to copy plot information from the printed map, which helped identify the total number of rice plots covering the extent of the sample mesh. Only plots that were either completely or partially inside the sample mesh were included in the listing process.

Figure 2: An Example of a 200 m x 200 m Mesh on Printed Map with Rice Plots Identified and Serially Numbered within the Mesh

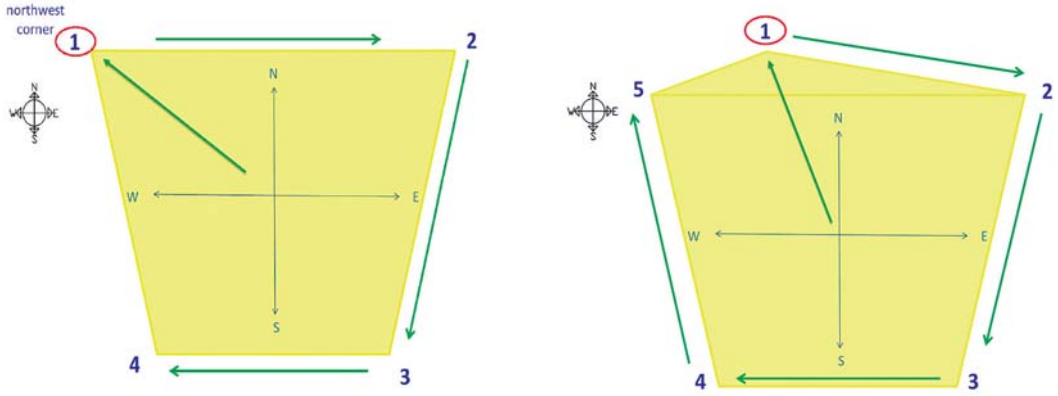
m = meter.
Source: Google Earth. 29 December 2015. Savannakhet, Lao PDR. 16°20'25.06" N, 104°57'14.97" E, eye alt 705m. http://www.earth.google.com (accessed 24 January 2017).

Systematic random sampling was then used to select a sample of four plots per mesh from the list of plots that met the selection criterion.[10] This involved calculating a sampling interval, which was used to systematically select the sample plots from the ordered list, following a random start. The selection of four plots was driven by the need to ensure sufficient sample size within a mesh to capture variability in rice yields across plots and budgetary constraints. For those sample meshes with four or less plots that were eligible for selection, crop cutting was done in all plots. If there were more than four plots within the mesh, crop cutting was implemented only on four randomly selected plots.

[10] Systematic sampling is one kind of probability sampling method wherein sample members from a larger population are selected according to a random starting point and a fixed, periodic interval known as the sampling interval.

At the third sampling stage, a random point was selected within each sample plot to identify a 2.5 m x 2.5 m crop-cutting subplot.[11] Crop cutting is a method wherein a small portion of a randomly selected plot, henceforth referred to as a subplot, is harvested, threshed, dried, and weighed to obtain objective yield estimates (Huddleston 1978). To identify the random point, the total number of corners was first listed for each randomly selected plot. The northwest-most corner was identified and labeled as corner number 1. Going clockwise, the remaining corners were numbered as illustrated in Figure 3.

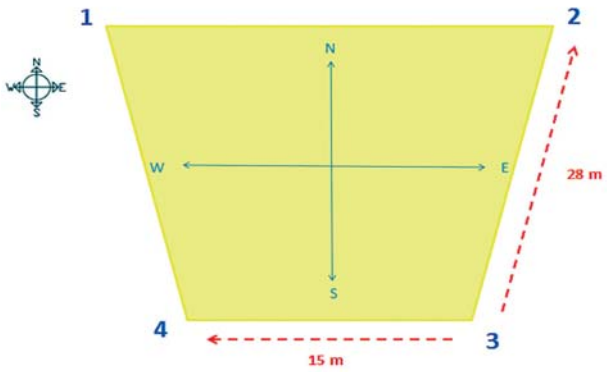
Figure 3: Numbering of Corners of Randomly Selected Plots



Source: Adapted from the presentation during the crop-cutting training in the pilot countries in 2017.

A random number table was used to select the corner where the field staff will start the random selection process to identify the subplot for the crop-cutting activity. The distance of the two sides along the selected random corner was measured with a measuring tape and/or with the GPS instrument. The longer side and the shorter side were identified and recorded as shown in Figure 4.

Figure 4: Measuring of Adjacent Sides of the Selected Corner



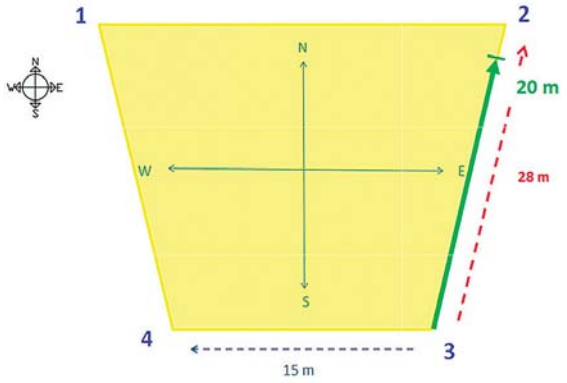
m = meter.
Source: Adapted from the presentation during the crop-cutting training in the pilot countries in 2017.

[11] In this study, whether rice was planted on a plot with the intention of harvesting in the rainy season of 2015 was the selection criterion since the objective was to only obtain estimates for one season using crop-cutting techniques.

After taking the measurements of the two adjacent sides of the random corner, the field staff took the bearing from the starting corner down the longer side and walked a random distance (determined from a random number table) which is less than the length of the longer side but in the direction of the longer side (Figure 5).

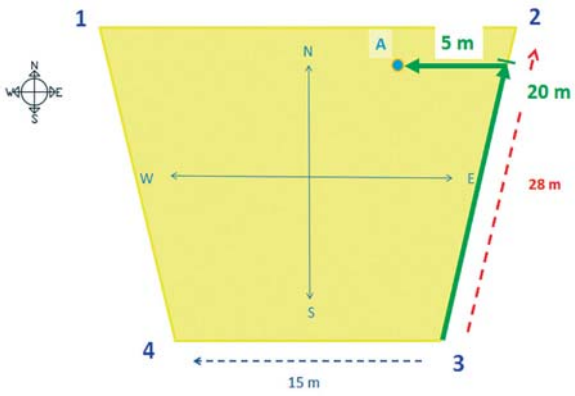
The field staff then entered the field in the direction parallel to the shorter side. The next random number from the second random number table that is lower than the length of the shorter side was chosen. This number corresponds to the length in meters that should be traversed inside the plot and parallel to the shorter side (Figure 6).

Figure 5: Plotting of Randomly Selected Distance from the Longer Arm



m = meter.
Source: Adapted from the presentation during the crop-cutting training in the pilot countries in 2017.

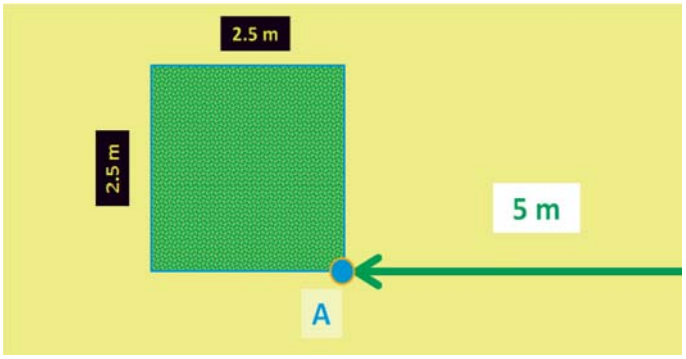
Figure 6: Plotting of Randomly Selected Distance from the Shorter Arm



m = meter.
Source: Adapted from the presentation during the crop-cutting training in the pilot countries in 2017.

Walking the random distance parallel to the short side and into the plot, the enumerator arrived at the starting point of the crop-cutting subplot presented in Figure 7. The 2.5 m x 2.5 m crop-cutting frame was placed at the upper right-hand corner of the starting point of the crop-cutting subplot. Wooden stakes were placed on the ground to keep the frame firm. Crop cutting was implemented on the subplot which involved harvesting all the paddy rice within. GPS measurements were also recorded for the starting point of the crop-cutting subplot. The harvested paddy was threshed, dried, and weighed for measuring the yield.

Figure 7: Identifying the Location of the 2.5 m x 2.5 m Subplot



m = meter.
Source: Adapted from the presentation during the crop-cutting training in the pilot countries in 2017.

The questionnaires were administered on paper in the countries' local language and were subsequently returned to the headquarters where the data entry took place.

B. Plot Level Paddy Rice Area Estimation

One of the most important components of the sample weighting procedures and the estimation of total production of rice paddy is the measurement of area planted in rice in sample plots within each sample mesh. In this study, two sources of objective measurements for the area of the sample rice plots were used:

(i) **Unmodified track data**. Unmodified tracks are based on the boundaries of the plot recorded by the enumerators using a handheld GPS navigation device. Field staff walked along the boundaries of each sample mesh to record its actual location and size (Figure 8). These track files were used to estimate the area of the plot and its intersection within the mesh.

(ii) **Modified track data**. GPS track files recorded were modified to correct the plotting of boundaries using Google Earth Pro (Figure 9). The modification was done because several issues were encountered during fieldwork that could influence the accuracy of the GPS tracks recorded. Boundaries of plots located on rough terrains were difficult to walk on. Some of the sampled plots were also located in areas with various obstructions which made passing through the boundaries too dangerous for the field staff. In such cases, field staff were asked to identify all obstructions and draw the correct boundaries of the plot on the printed map which was factored into the digitization process.

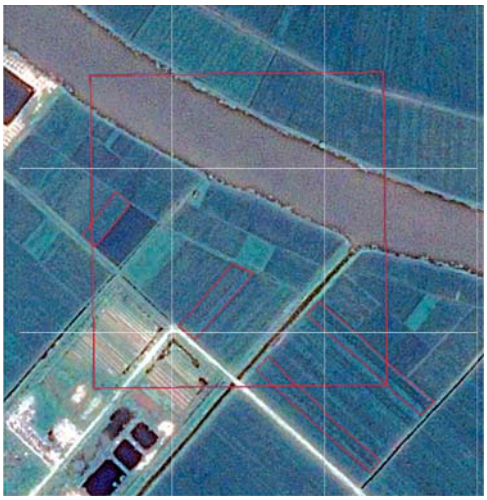
A GIS software named QGIS was used for data processing.[12] Plot area and intersection estimates based on both unmodified and modified tracks were derived using GIS techniques to facilitate a comparison of the estimates of total area planted in paddy rice from the two different sources, to assess their relative accuracy, and selection for final estimation procedures.

[12] QGIS (previously known as "Quantum GIS") is a cross-platform free and open-source desktop geographic information system (GIS) application that provides data viewing, editing, and analysis capabilities. QGIS version 2.14 (Essen) was used for related exercises under R-CDTA 8369: Innovative Data Collection Methods for Agricultural and Rural Statistics.

Figure 8: Unmodified Global Positioning System Track Data on Google Earth

Source: Google Earth. 26 August 2016. Thai Binh, Viet Nam. 20°35'29.11" N, 106°24'24.66" E, eye alt 562m. http://www.earth.google.com (accessed 13 October 2016).

Figure 9: Modified Track Data on Google Earth



Source: Google Earth. 26 August 2016. Thai Binh, Viet Nam. 20°35'29.11" N, 106°24'24.66" E, eye alt 562m. http://www.earth.google.com (accessed 13 October 2016).

IV. WEIGHTING, SAMPLING ERROR, AND DESIGN EFFECTS

A. Weighting Procedure

The weight for each sampling unit is based on the inverse of its overall probability of selection, taking into account all sampling stages. In the case of the sample meshes in each stratum, the probability of selection would be the following:

$$p_{1h} = \frac{n_h}{N_h} \tag{1}$$

where:

p_{1h} = first stage probability of selection of each sample mesh in stratum *h*,
n_h = number of sample meshes selected and visited in stratum *h*,
N_h = total number of meshes in frame for stratum *h*.

The second stage probability of selection for the sample plots is calculated as the number of sample plots selected in the mesh (generally four) divided by the total number of plots with rice planted during the season totally or partly within the mesh. This probability can be expressed as follows:

$$p_{2hi} = \frac{n_{hi}}{N_{hi}} \tag{2}$$

where:

p_{2hi} = second stage probability of selection of each sample plot with rice within the *i*-th sample mesh in stratum *h*,
n_{hi} = number of sample plots selected for crop cutting in the *i*-th sample mesh in stratum *h* (generally equal to four),
N_{hi} = total number of plots with rice planted this season that are at least partly located within the boundaries of the *i*-th sample mesh in stratum *h* (including those that had already been harvested and those where the crop was lost for any reason).

Although the plots for which the rice had already been harvested or where rice was planted but lost pre-harvest are excluded from the second stage selection of plots (since no crop cutting could be conducted), these plots are still represented in the estimates and included in the denominator of this probability.

At the third stage, one 2.5 m x 2.5 m subplot is selected within each of the maximum of 4 sample plots in the mesh for the rice paddy crop cutting. In this case the probability will depend on the total area of the plot. Therefore, the third stage probability of selection can be calculated as follows:

$$p_{3hij} = \frac{6.25m^2}{a_{hij}} \tag{3}$$

where:

p_{3hij} = third stage probability of selection of the sample subplot in the *j*-th sample plot with rice within the *i*-th sample mesh in stratum *h*,
a_{hij} = total area (in m^2) of the *j*-th sample plot with rice within the *i*-th sample mesh in stratum *h*.

The weight for the sample mesh would be calculated as the inverse of the first stage probability of selection as follows:

$$W_{1h} = \frac{1}{p_{1h}} = \frac{N_h}{n_h} \tag{4}$$

where:

W_{1h} = basic weight for each sample mesh in stratum *h*.

The weight for the sample plots in each sample mesh was based on the inverse of the overall probability of selection, which involved both the first and second stage probabilities of selection. In the case where part of the sample plot is located outside the mesh boundaries, it was necessary to adjust this weight based on the proportion of the plot that is inside the mesh boundaries. The reason for this adjustment is that the plot can be selected if either of the two adjacent meshes is selected. If we did not adjust the weight, we would be overestimating the weighted total area at the stratum level. The general expression for the weight of a sample plot can be defined as follows:

$$W_{2hij} = \frac{1}{p_{1h} \times p_{2hi}} \times \frac{a'_{hij}}{a_{hij}} = \frac{N_h}{n_h} \times \frac{N_{hi}}{n_{hi}} \times \frac{a'_{hij}}{a_{hij}} \tag{5}$$

where:

W_{2hij} = basic weight for the *j*-th sample plot with rice within the *i*-th sample mesh in stratum *h*,

a'_{hij} = area of the *j*-th sample plot with rice inside the boundaries of the *i*-th sample mesh in stratum *h*.[13]

When the area of the sample plot is completely within the boundaries of the sample mesh, $a'_{hij} = a_{hij}$, so the last component of this weight would be equal to 1. It should be noted that the weight of the plot expressed above is applied to the data for the entire plot, including the area outside the sample mesh. The formula for the total area planted in rice can be expressed as follows:

$$\hat{A} = \sum_h \sum_i \sum_j W_{2hij} \times a_{hij} = \sum_h \sum_i \sum_j \frac{N_h}{n_h} \times \frac{N_{hi}}{n_{hi}} \times \frac{a'_{hij}}{a_{hij}} \times a_{hij} = \sum_h \sum_i \sum_j \frac{N_h}{n_h} \times \frac{N_{hi}}{n_{hi}} \times a'_{hij} \quad (6)$$

Since we also had information on the area of rice planted in the part of the plot inside the mesh (a'_{hij}), we can use a simpler expression to calculate the total area of rice. In the case of the sample subplot in each sample plot, the weight will include probability components from all three sampling stages. The weight for each subplot can be expressed as follows:

$$W_{3hij} = \frac{1}{p_{1h} \times p_{2hi} \times p_{3hij}} = \frac{N_h}{n_h} \times \frac{N_{hi}}{n_{hi}} \times \frac{a'_{hij}}{a_{hij}} \times \frac{a_{hij}}{6.25m^2} = \frac{N_h}{n_h} \times \frac{N_{hi}}{n_{hi}} \times \frac{a'_{hij}}{6.25m^2} \quad (7)$$

where:

W_{3hij} = basic weight for the sample subplot in the *j*-th sample plot within the *i*-th sample mesh in stratum *h*.

B. Calculation of Sampling Errors and Design Effects

To evaluate the results, it is important to examine the estimated accuracy of the survey data. In addition to presenting tables with the calculated sampling errors and confidence intervals for the most important survey estimates, we also provide some theoretical reasons for the different sources of nonsampling error, while noting that nonsampling errors are near impossible to detect and eliminate in any study.[14]

The standard error (SE), or square root of the variance, is used to measure the sampling error, although it may also include a small variable part of the nonsampling error. The coefficient of variation (CV), 95% confidence interval, the design effect (DEFF), and the number of observations were also calculated. The design effect is defined as the variance (square of the sampling error) of an estimate based on the actual stratified multistage sample design divided by the corresponding variance from a simple random sample of the same size. It is a measure of the relative efficiency of the sample design. A design effect much larger than 1 indicates a large clustering effect, with a high correlation of the variables (such as plot area) within the sample meshes.

---

[13] The ratio $\frac{a'_{hij}}{a_{hij}}$ is also known as the weighted segment estimator because the estimator is based on the ratio of the area of the plot in the mesh to the land area in the entire plot.

[14] Nonsampling error is caused by factors other than those related to sample selection. It refers to the presence of any factor, whether systemic or random, that results in the data values not accurately reflecting the "true" value for the population. This includes coverage, nonresponse, response, interviewer, and processing errors.

The variance estimator of a total can be expressed as follows:

$$V(\hat{Y}) = \sum_{h=1}^{L} \left[\frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left(\hat{Y}_{hi} - \frac{\hat{Y}_h}{n_h} \right)^2 \right] \quad (8)$$

where:

$V(\hat{Y})$ = variance of the estimate of the total for variable *y* such as total area

$\hat{Y}_{hi} = \sum_{j=1}^{m_h} W'_{hi}\, y_{hij}$ = weighted total of variable *y* (such as area) for the *i*-th sample mesh

W'_{hi} = weight of sample plots in the *i*-th sample mesh in stratum *h*

Y_{hij} = value of variable *y* for the *j*-th sample plot in the *i*-th sample mesh in stratum *h*

$\hat{Y}_h = \sum_{i=1}^{n_h} \hat{Y}_{hi}$ = weighted total of variable *y* for stratum *h*

$\hat{Y} = \sum_h \hat{Y}_h$ = weighted estimate of the total for variable *y*

The linearized Taylor-series variance estimator of a ratio used can be expressed as follows:

$$V(\hat{R}) = \frac{1}{\hat{X}^2} \left[V(\hat{Y}) + \hat{R}^2 V(\hat{X}) - 2\hat{R} COV(\hat{X}, \hat{Y}) \right] \quad (9)$$

where:

$$COV(\hat{X}, \hat{Y}) = \sum_{h=1}^{L} \left[\frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left(\hat{X}_{hi} - \frac{\hat{X}_h}{n_h} \right) \left(\hat{Y}_{hi} - \frac{\hat{Y}_h}{n_h} \right) \right] \quad (10)$$

Here, $V(\hat{Y})$ and $V(\hat{X})$ are calculated according to the formula for the variance of a total, while $\hat{R} = \frac{\hat{Y}}{\hat{X}}$ is the estimate of a ratio, where $\hat{Y}$ and $\hat{X}$ are weighted totals for the variables *y* and *x*.

V. RESULTS

A. Comparative Analysis of Direct Estimates of Total Area Planted in Rice

Two sets of weights were calculated for the sample subplots based on each source of area measurement—unmodified and modified data. These weights were used to produce alternative direct estimates of the total area planted in rice alongside calculating corresponding SE, 95% confidence intervals and design effects. Table 3 presents the estimates of the total area planted in rice paddy in m^2 for each province by stratum and corresponding estimates of the level of precision, using the weights based on the area of the rice plots inside the sample mesh from the unmodified track data. Table 4 presents the corresponding results using the weights based on the modified track data. Figure 10 shows a comparison of the estimates of total area planted in rice by stratum for each province based on the two alternative area estimation procedures.

Table 3: Standard Error, Coefficient of Variation, and Design Effects of Estimates of Total Area Planted in Rice Paddy Based on Area of Sample Plots from Unmodified Track Data

Domain/Stratum	Area (m^2)	SE	CV	95% Confidence Interval		DEFF	No. of Observations
				Lower	Upper		
Savannakhet	2,700,047,316	509,013,893	0.189	1,685,814,995	3,714,279,638	6.25	136
IRRI+GlobCover	1,092,874,253	112,474,296	0.103	868,764,326	1,316,984,181	0.38	105
IRRI	44,413,015	12,865,940	0.290	18,777,070	70,048,961	0.08	16
GlobCover	804,851,923	203,242,057	0.253	399,883,285	1,209,820,561	1.73	13
Other	757,908,125	452,738,125	0.597	0	1,660,008,531	10.98	2
Ang Thong	292,337,345	40,269,296	0.138	210,408,846	374,265,843	1.69	104
IRRI+GlobCover	274,834,018	39,838,268	0.145	193,782,453	355,885,582	1.62	82
IRRI	2,779,591	976,108	0.351	793,684	4,765,498	0.03	8
GlobCover	14,723,736	5,794,467	0.394	2,934,805	26,512,667	0.71	14
Thai Binh	474,230,049	45,807,978	0.097	382,631,336	565,828,761	4.34	256
IRRI+GlobCover	446,207,926	45,109,146	0.101	356,006,614	536,409,237	3.97	220
IRRI	110,039	19,651	0.179	70,744	149,334	0	8
GlobCover	27,912,084	7,970,911	0.286	11,973,262	43,850,906	0.92	28

CV = coefficient of variation, DEFF = design effects, m^2 = square meter, SE = sampling error.
Notes: IRRI refers to meshes identified as paddy area by the International Rice Research Institute (IRRI) rice maps. GlobCover refers to meshes identified as paddy area by the European Space Agency (ESA) GlobCover land use maps. IRRI+GlobCover refers to meshes identified as paddy area by both IRRI and ESA area maps.
Source: Authors' estimates.

Table 4: Standard Error, Coefficient of Variation, and Design Effects of Estimates of Total Area Planted in Rice Paddy Based on Area of Sample Plots from Modified Track Data

Domain/Stratum	Area (m^2)	SE	CV	95% Confidence Interval		DEFF	No. of Observations
				Lower	Upper		
Savannakhet	2,496,699,709	410,394,222	0.164	1,678,971,372	3,314,428,047	4.91	136
IRRI+GlobCover	1,086,335,396	108,016,428	0.099	871,107,964	1,301,562,827	0.37	105
IRRI	44,287,871	13,014,555	0.294	18,355,804	70,219,939	0.08	16
GlobCover	714,442,692	189,790,467	0.266	336,276,931	1,092,608,453	1.98	13
Other	651,633,750	347,226,250	0.533	0	1,343,497,156	9.35	2
Ang Thong	299,114,441	41,045,835	0.137	215,606,062	382,622,821	1.65	104
IRRI+GlobCover	278,414,050	40,259,423	0.145	196,505,638	360,322,462	1.56	82
IRRI	2,567,025	982,144	0.383	568,838	4,565,213	0.04	8
GlobCover	18,133,366	7,935,667	0.438	1,988,130	34,278,603	0.87	14
Thai Binh	484,750,036	45,200,548	0.093	394,365,955	575,134,118	4.89	256
IRRI+GlobCover	456,083,074	44,435,139	0.097	367,229,522	544,936,626	4.36	220
IRRI	109,143	14,156	0.130	80,837	137,450	0	8
GlobCover	28,557,819	8,282,979	0.290	11,994,980	45,120,658	1.02	28

CV = coefficient of variation, DEFF = design effects, m^2 = square meter, SE = sampling error.
Notes: IRRI refers to meshes identified as paddy area by the International Rice Research Institute (IRRI) rice maps. GlobCover refers to meshes identified as paddy area by the European Space Agency (ESA) GlobCover land use maps. IRRI+GlobCover refers to meshes identified as paddy area by both IRRI and ESA area maps.
Source: Authors' estimates.

In both Ang Thong and Thai Binh, roughly 93% of the estimate of total rice area is from the *IRRI+GlobCover* stratum in both unmodified and modified track files. The estimates for this stratum have higher design effects, which are mostly measuring the clustering effects due to the similarity of the plot areas within a mesh. The results differ in Savannakhet where only about 40% of the estimate of total rice area comes from the *IRRI+GlobCover* stratum. On the other hand, more than 26% of the weighted area comes from two sample meshes in the *Other* stratum, which has a much lower sampling rate and thus, a much higher weight. Since the *Other* stratum was defined as areas with a very small probability of growing rice, this indicates a problem with the efficiency of the stratification of the sampling frame, either in discriminating the rice found in this stratum from the satellite images, or because of more recent rice planting activities after the satellite images were generated. Also, the estimate of total paddy area for Savannakhet is higher for the unmodified track area data, while the modified track area data produced higher estimates for both Ang Thong and Thai Binh.[15]

It is also important to examine the CVs, SEs, and confidence intervals for the weighted estimates of total paddy area based on unmodified and modified tracks for each province to compare the level of precision. The highest CV was calculated from estimates of total paddy area for Savannakhet—18.9% for unmodified and 16.4% for modified. This was followed by the CVs of estimates for Ang Thong (greater than 13%) and Thai Binh (9%) for both unmodified and modified tracks. These resulted in relatively wide confidence intervals for the estimates for all pilot provinces. One reason for this is the variability in the size of the sample plots selected randomly within each mesh.[16]

Figure 10 also shows that the estimate of total area of rice planted based on the modified track area data is higher than the corresponding estimate based on the unmodified track data for Ang Thong and Thai Binh. However, given the wide confidence intervals, the differences are not statistically significant.

Given the greater quality control involved in measuring the modified track area, the values are considered more accurate than the unmodified track data. For Thai Binh, the assumption of overall less variability in the modified data is verified in Table 5, which shows estimates of the statistical parameters for the unweighted data for rice area of plot inside the mesh from each source. The unmodified track data has a higher standard deviation and a higher maximum value than the corresponding area values in the modified track data. Table 5 also shows a higher unweighted mean area value for the modified track data for Thai Binh, which explains the higher estimates of total area in rice for the data from this source. Hence, the subplot weights based on the modified area of rice plots inside the mesh are used for the remainder of this analysis.[17]

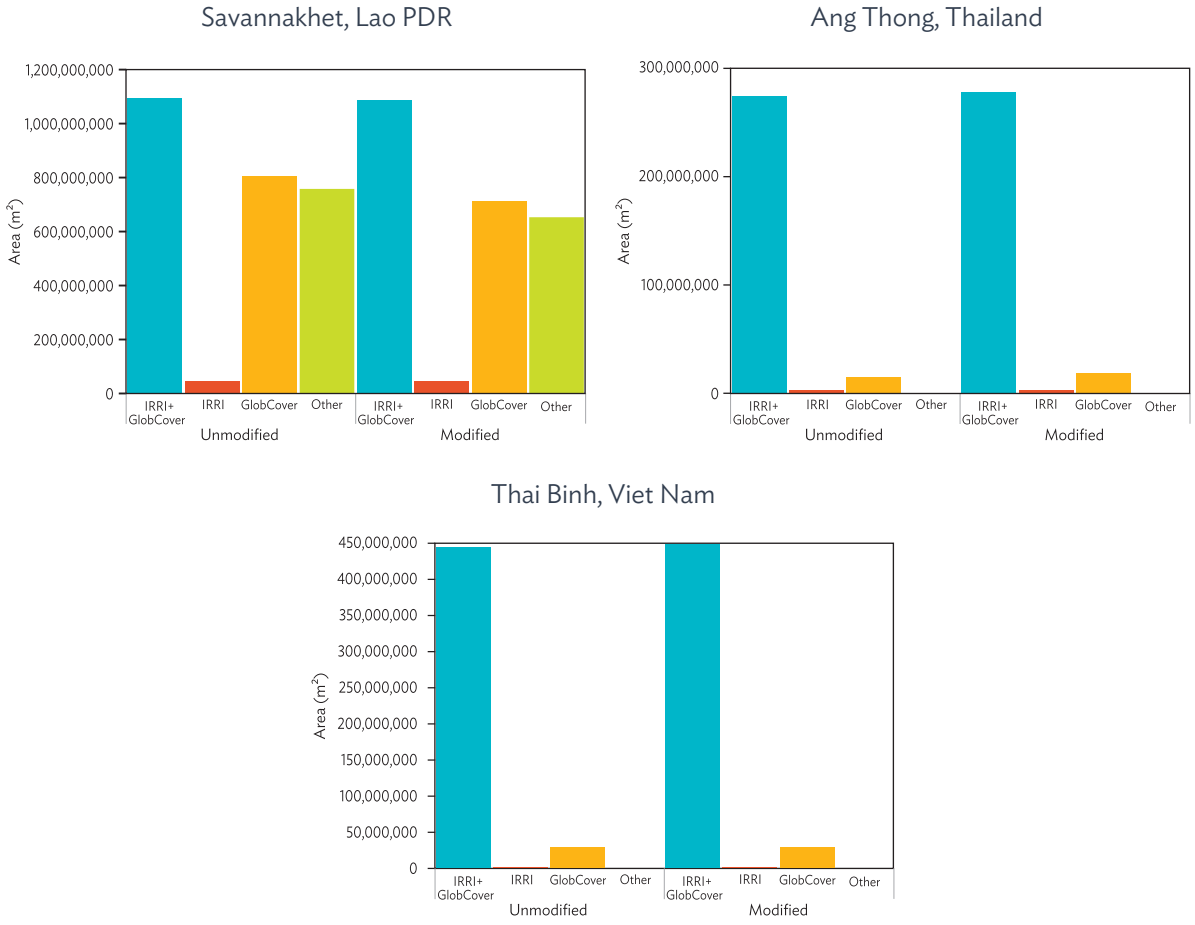
---

[15] It should be noted that rice plots in Savannakhet were relatively larger in size than those in Thai Binh. Thus, there were fewer plots within the meshes in Savannakhet compared to Thai Binh.

[16] As will be indicated in the next section, this also results in high CVs for the estimates of total production of rice paddy.

[17] Tabulations of the average yield and total production of rice paddy using both sets of weights were also calculated, and as expected, the weights based on the modified track data resulted in slightly higher estimates of total production. However, the estimate of average yield per m^2 was the same using both sets of weights.

Figure 10: Paddy Area Estimates in Savannakhet, Ang Thong, and Thai Binh Based on Data from Unmodified and Modified Track Files



Lao PDR = Lao People's Democratic Republic, m^2 = square meter.
Notes: IRRI refers to meshes identified as paddy area by the International Rice Research Institute (IRRI) rice maps. GlobCover refers to meshes identified as paddy area by the European Space Agency (ESA) GlobCover land use maps. IRRI+GlobCover refers to meshes identified as paddy area by both IRRI and ESA area maps.
Source: Authors' estimates.

Table 5: Estimates of Statistical Parameters for the Unweighted Paddy Area Data by Source

Pilot Area	Source of Data	Mean (m^2)	Standard Deviation	Minimum Value	Maximum Value
Savannakhet	Unmodified track data	5,758.35	5,465.67	52.97	27,488.38
	Modified track data	5,676.47	5,365.19	7.83	28,260.85
Ang Thong	Unmodified track data	3,444.54	3,442.93	0	17,803.71
	Modified track data	3,497.89	3,452.85	0	17,086.20
Thai Binh	Unmodified track data	720.02	522.56	39.70	3,190.60
	Modified track data	729.39	497.79	31.00	2,743.70

m^2 = square meter.
Source: Authors' estimates.

B. Yield Estimates Derived from Crop-Cutting Exercise

Table 6 shows the weighted estimates of the average kilogram (kg) of rice paddy harvested in the 6.25 m^2 subplots and the corresponding sampling error, while Table 7 shows the corresponding average yield in kg/m^2. The estimate of the average yield is based on the weighted mean of the subplot crop-cutting data.

The CV and design effect for the estimates in Tables 6 and 7 are the same, since these estimates only vary by a constant value. The relative confidence interval is narrower than the corresponding estimates of total area and production of paddy rice. This estimate of average yield is a type of ratio, and the correlation between the numerator and denominator of the ratio reduces the corresponding sampling error.

Figure 11 presents the average yield per pilot province graphically in tons/ha and compares it to yield estimates obtained from official data sources. It is interesting to note that the crop-cutting yields are about the same in the case of Ang Thong and Thai Binh, with the minor differences emanating from our specification of moisture content to 12% in this study. However, for Savannakhet, the yield from crop cutting is nearly half of the official figure, warranting a closer look at the methodology implemented to produce official statistics.

Table 6: Estimate of Mean Yield of Rice Paddy per Subplot

Domain	Estimate (kg/6.25 m^2)	SE	CV	95% Confidence Interval		DEFF	No. of Observations
				Lower	Upper		
Savannakhet	1.20	0.07	0.058	1.06	1.34	4.36	136
IRRI+GlobCover	1.34	0.07	0.052	1.20	1.48	1.47	105
IRRI	1.22	0.15	0.122	0.92	1.51	0.45	16
GlobCover	1.17	0.16	0.141	0.84	1.50	6.48	13
Other	1.01	0.05	0.051	0.90	1.11	25.23	2
Ang Thong	2.22	0.12	0.053	1.98	2.46	2.57	103
IRRI+GlobCover	2.22	0.12	0.055	1.97	2.46	2.67	81
IRRI	1.88	0.28	0.147	1.32	2.44	0.23	8
GlobCover	2.28	0.54	0.237	1.18	3.38	1.92	14
Thai Binh	3.36	0.08	0.025	3.19	3.53	4.12	253
RRI+GlobCover	3.35	0.09	0.026	3.17	3.53	4.29	219
IRRI	2.50	0.45	0.181	1.59	3.40	0.03	8
GlobCover	3.54	0.24	0.069	3.05	4.03	1.97	26

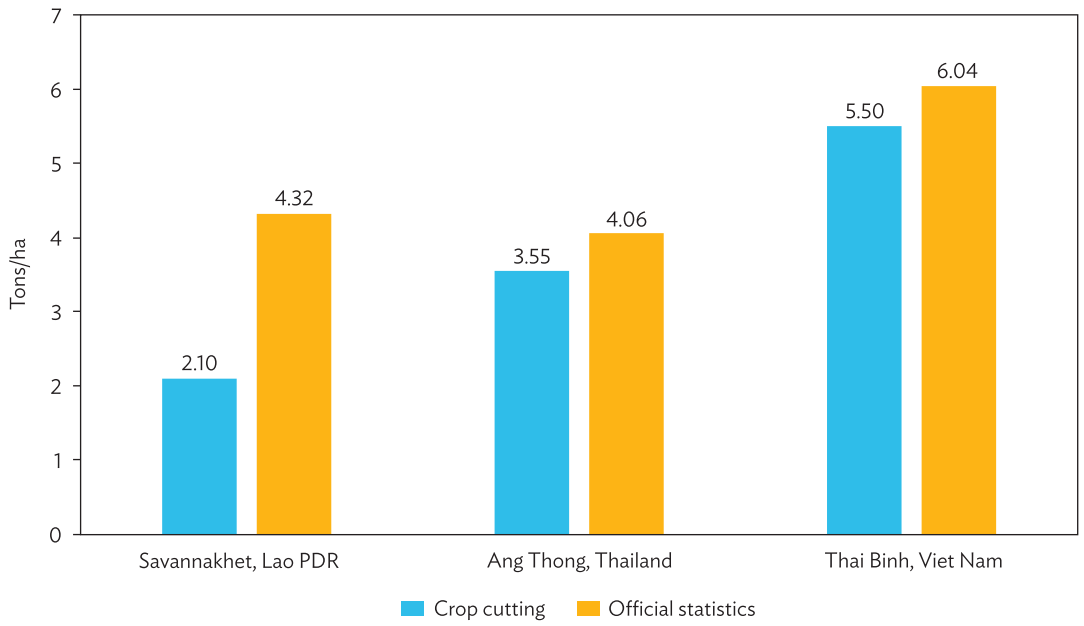
CV = coefficient of variation, DEFF = design effects, kg = kilogram, m^2 = square meter, SE = sampling error.
Notes: IRRI refers to meshes identified as paddy area by the International Rice Research Institute (IRRI) rice maps. GlobCover refers to meshes identified as paddy area by the European Space Agency (ESA) GlobCover land use maps. IRRI+GlobCover refers to meshes identified as paddy area by both IRRI and ESA area maps.
Source: Authors' estimates.

Table 7: Estimate of Mean Yield

Domain	Yield, (kg/m^2)	SE	CV	95% Confidence Interval		DEFF	No. of Observations
				Lower	Upper		
Savannakhet	0.19	0.01	0.058	0.17	0.21	4.36	136
IRRI+GlobCover	0.21	0.01	0.052	0.19	0.24	1.47	105
IRRI	0.19	0.02	0.122	0.15	0.24	0.45	16
GlobCover	0.19	0.03	0.141	0.13	0.24	6.48	13
Other	0.16	0.01	0.051	0.14	0.18	25.23	2
Ang Thong	0.35	0.02	0.053	0.32	0.39	2.57	103
IRRI+GlobCover	0.35	0.02	0.055	0.32	0.39	2.67	81
IRRI	0.30	0.04	0.147	0.21	0.39	0.23	8
GlobCover	0.36	0.09	0.237	0.19	0.54	1.92	14
Thai Binh	0.54	0.01	0.025	0.51	0.56	4.12	253
IRRI+GlobCover	0.54	0.01	0.026	0.51	0.56	4.29	219
IRRI	0.40	0.07	0.181	0.25	0.54	0.03	8
GlobCover	0.57	0.04	0.069	0.49	0.64	1.97	26

CV = coefficient of variation, DEFF = design effects, kg/m^2 = kilogram per square meter, SE = sampling error.
Notes: IRRI refers to meshes identified as paddy area by the International Rice Research Institute (IRRI) rice maps. GlobCover refers to meshes identified as paddy area by the European Space Agency (ESA) GlobCover land use maps. IRRI+GlobCover refers to meshes identified as paddy area by both IRRI and ESA area maps.
Source: Authors' estimates.

Figure 11: Average Yield per Pilot Area



ha = hectare, Lao PDR = Lao People's Democratic Republic, tons = metric tons.
Source: Authors' estimates.

C. Direct Estimates of Total Production of Rice Paddy

The direct estimation of total production (kg) and the average yield (kg/m^2) of harvested paddy are derived from crop-cutting data for the sample of 2.5 m x 2.5 m (i.e., 6.25 m^2) subplots by multiplying data on harvested paddy in each subplot by the corresponding subplot weights used for the estimation of total area. The direct estimates of the total production of harvested paddy using the weights based on the area of the rice plots inside the mesh from the modified track data are presented in Table 8, with corresponding estimates of the measures of precision. The components of the estimates of total production by stratum for each province are shown in Figure 12.

Table 8 shows relatively high CVs for the direct estimate of the total production of rice for Savannakhet, Ang Thong, and Thai Binh, resulting in corresponding wide confidence intervals. The main source of this discrepancy is due to the variability in the estimate of total area planted in rice paddy.

Table 8: Direct Estimates of Total Production of Rice Paddy

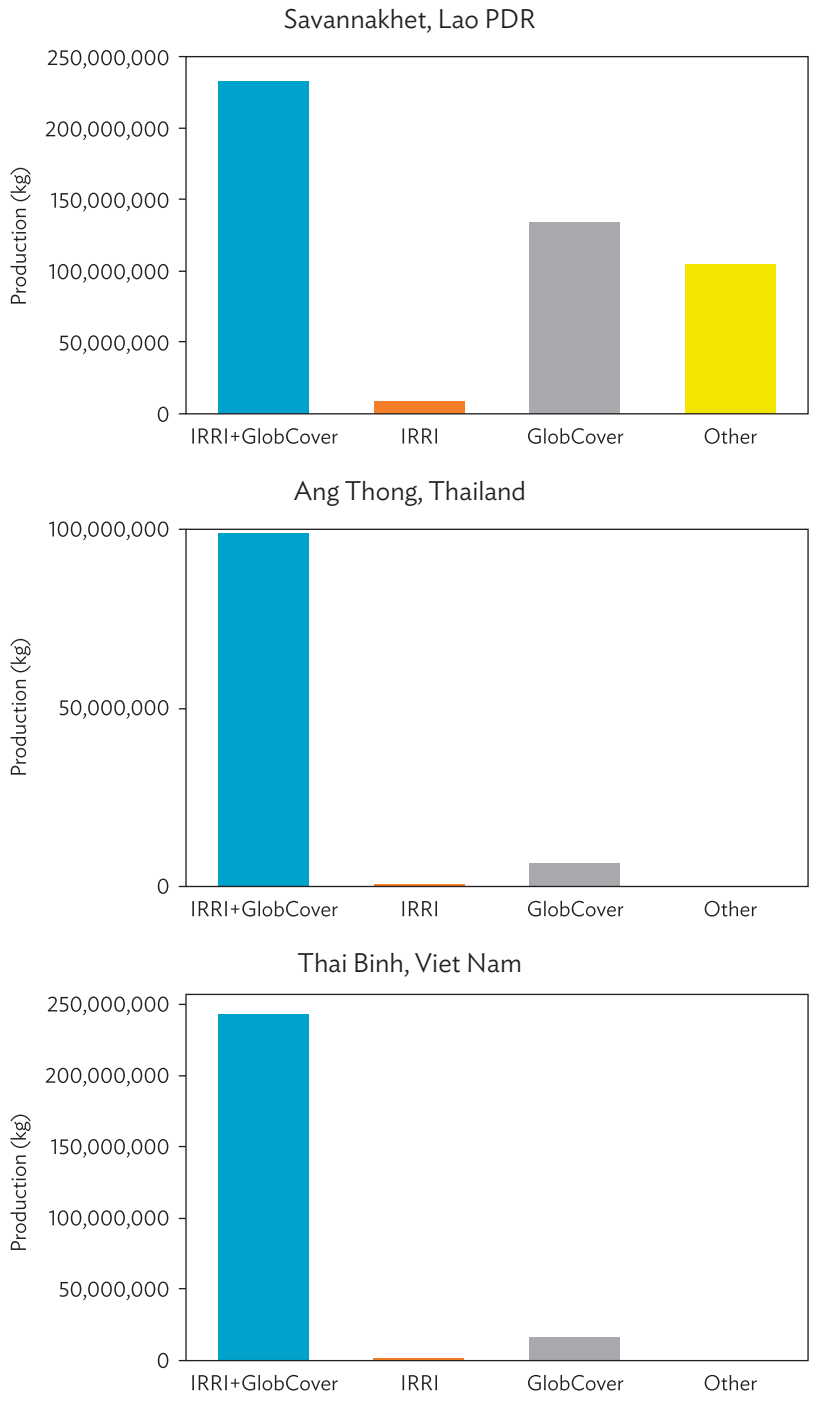
Domain/Stratum	Production (kg)	SE	CV	95% Confidence Interval		DEFF	No. of Observations
				Lower	Upper		
Savannakhet	480,210,025	70,689,811	0.147	339,357,502	621,062,547	27.78	136
IRRI+GlobCover	232,891,720	28,133,072	0.121	176,835,350	288,948,089	1.27	105
IRRI	8,612,733	2,300,195	0.267	4,029,494	13,195,972	0.16	16
GlobCover	133,816,763	40,578,604	0.303	52,962,128	214,671,397	4.27	13
Other	104,888,810	50,533,806	0.482	4,198,002	205,579,617	11.01	2
Ang Thong	106,137,971	13,871,631	0.131	77,915,925	134,360,017	15.63	103
IRRI+GlobCover	98,753,252	13,644,497	0.138	70,993,313	126,513,191	10.31	81
IRRI	772,357	319,637	0.414	122,050	1,422,663	0.14	8
GlobCover	6,612,362	2,479,451	0.375	1,567,882	11,656,842	0.78	14
Thai Binh	260,670,296	27,265,098	0.105	206,150,363	315,190,230	72.01	253
IRRI+GlobCover	244,447,928	26,764,357	0.109	190,929,289	297,966,567	29.19	219
IRRI	43,599	2,245	0.051	39,111	48,087	0	8
GlobCover	16,178,769	5,201,422	0.321	5,777,883	26,579,656	1.57	26

CV = coefficient of variation, DEFF = design effects, kg = kilogram, SE = sampling error.
Notes: IRRI refers to meshes identified as paddy area by the International Rice Research Institute (IRRI) rice maps. GlobCover refers to meshes identified as paddy area by the European Space Agency (ESA) GlobCover land use maps. IRRI+GlobCover refers to meshes identified as paddy area by both IRRI and ESA area maps.
Source: Authors' estimates.

D. Obtaining Mesh Level Estimates of Total Area Planted in Rice from Modified Map

Direct estimates of the total area planted in rice from the sample plots have a relatively high coefficient of variation because of the variability in the sizes of the sample plots. However, mesh level information of total area planted may improve the precision of the estimates. To produce mesh level estimates for total area planted in rice, paper maps of sample meshes with recent Google Earth images were used. Enumerators filled out additional information to delineate rice area planted within each mesh on the paper maps, which were subsequently digitized using the ALIS methodology.

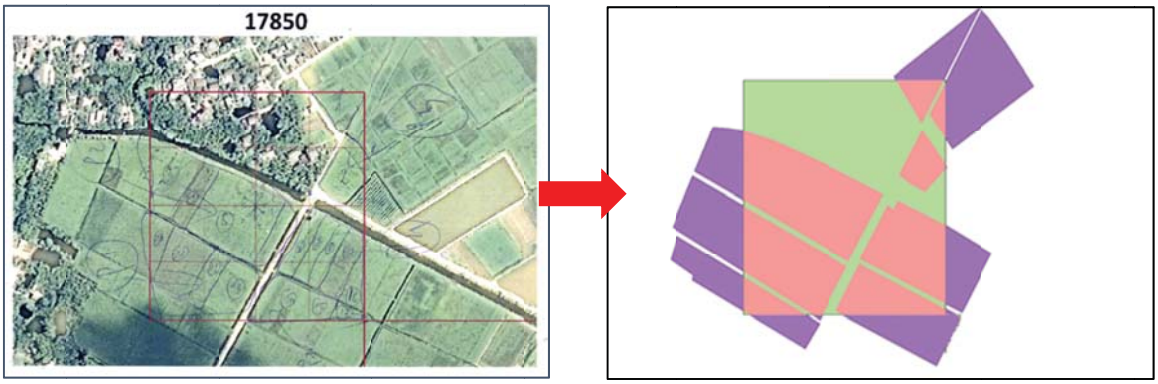
Figure 12: Total Production per Pilot Province by Stratum



kg = kilogram, Lao PDR = Lao People's Democratic Republic.
Notes: IRRI refers to meshes identified as paddy area by the International Rice Research Institute (IRRI) rice maps. GlobCover refers to meshes identified as paddy area by the European Space Agency (ESA) GlobCover land use maps. IRRI+GlobCover refers to meshes identified as paddy area by both IRRI and ESA area maps.
Source: Authors' estimates.

The ALIS methodology was originally developed under the Association of Southeast Asian Nations (ASEAN) Food Security Information System (AFSIS) initiative and provides for strict digitization guidelines to identify crop area planted using visual inspection at specific GPS locations for major crops such as rice, maize, cassava, sugarcane, and soybean.[18] It was first adopted and successfully implemented by AFSIS in Vientiane Province, Lao PDR; in Kandal Province, Cambodia; and in Nueva Ecija, Philippines.[19]

Figure 13: Mesh Information on a Printed Map



QGIS = quantum geographic information system.
Source: Based on the field validation of the Center for Informatics and Statistics, Ministry of Agriculture and Rural Development, Viet Nam and digitization using QGIS.

The area intersection between the mesh and the digitized rice fields were then derived and used to estimate the percentage of rice area or rice area rate within each sample mesh. Figure 13 shows a printed map with information on the different plots identified inside the mesh which was subsequently converted to a digital map.

The estimation of total area in rice was simplified by compiling a database of sample meshes, with summary information including (i) sample mesh ID, (ii) sample mesh weight, and (iii) total area of rice planted inside the boundaries of the sample mesh for all plots (including plots that had already been harvested and plots where rice was planted but the harvest was completely lost for any reason) from digitized maps.

The total area planted in rice is calculated using the following formula:

$$\hat{A} = \sum_{h} \hat{A}_h = \sum_{h} \sum_{i \in h} W_{1h} A_{hi} \tag{11}$$

[18] AFSIS is one of the initiatives supported by the ASEAN+3 Cooperation Framework. The AFSIS Project aimed to strengthen food security in the region through the systematic collection, analysis, and dissemination of food security-related information.

[19] The implementation of the ALIS Project in the Philippines was supported by the Asian Development Bank under the regional policy advocacy technical assistance (R-PATA) 8029. For detailed discussion on ALIS: https://www.adb.org/publications/results-methodological-studies-agricultural-and-rural-statistics.

where:

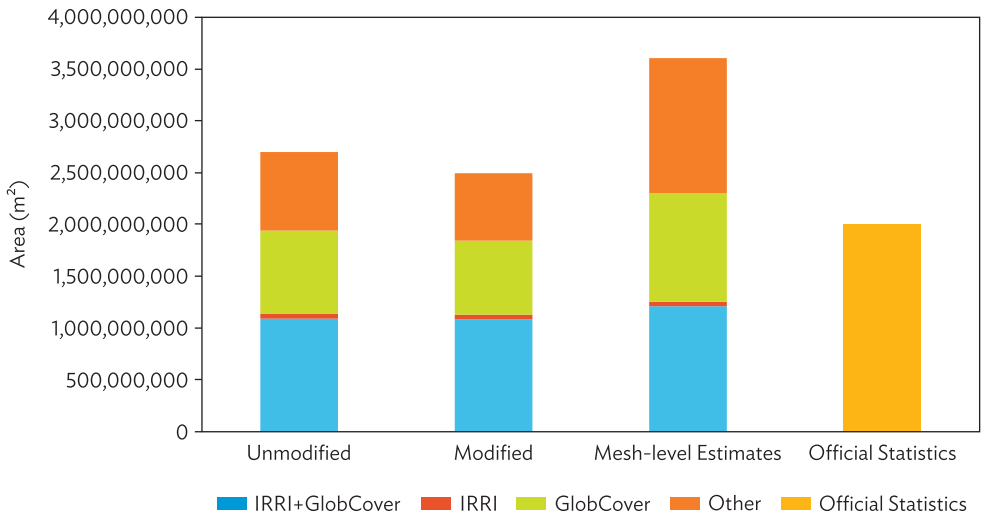
$\hat{A}$ = weighted estimate of total area planted in rice in the province
$\hat{A}_h$ = weighted estimate of total area planted in rice in stratum h
W_{1h} = basic weight for each sample mesh in stratum h
A_{hi} = total area planted in rice within the boundaries of the i-th sample mesh in stratum h, from the digitized map

The variable A_{hi} corresponds to the third variable listed above for the mesh-level database. It should be noted that the measurement of area A_{hi} in equation 11 is at the mesh level, whereas the measurement of area a_{hij} in equation 6 is at the plot level.

E. Comparison of Area Estimates from Different Data Sources

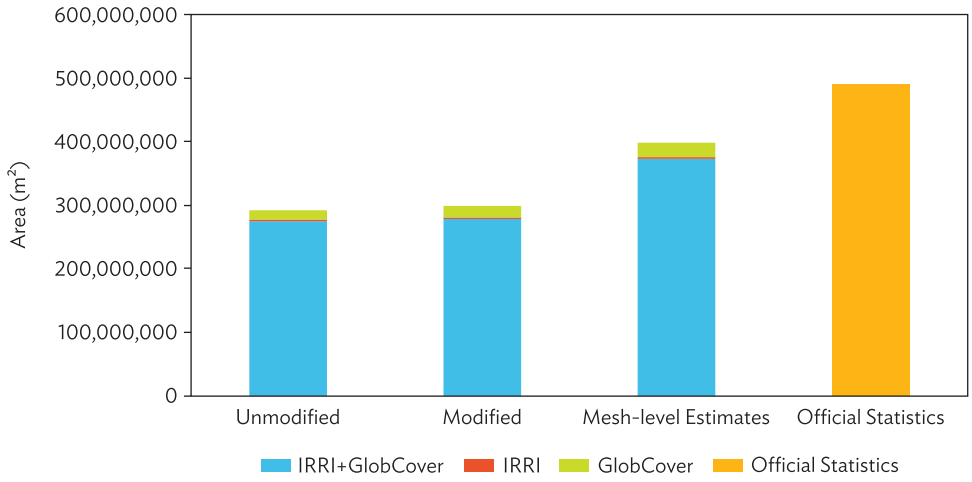
Figures 14, 15, and 16 present the weighted estimate of total paddy planted area using digitized Google Earth images at the mesh level compared to area estimates from other sources.

Figure 14: Estimates of Total Planted Area in Savannakhet from Different Data Sources



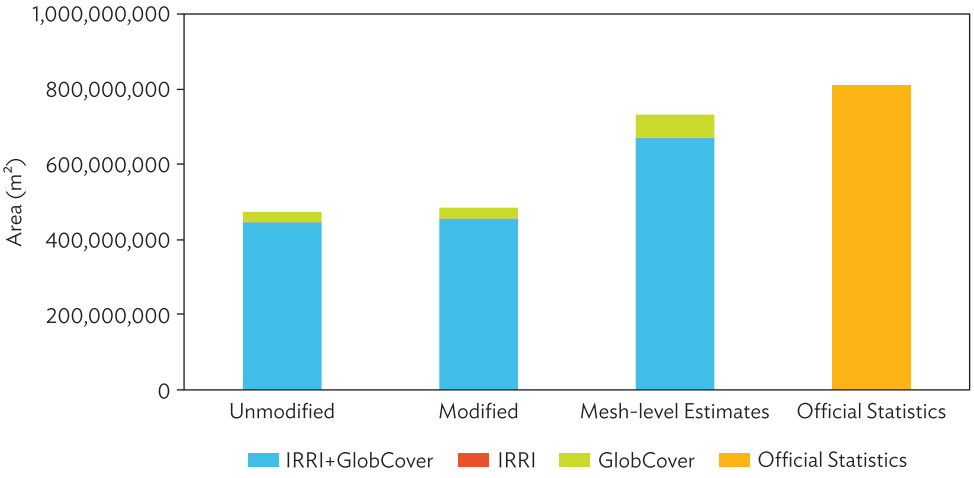
m^2 = square meter.
Notes: IRRI refers to meshes identified as paddy area by the International Rice Research Institute (IRRI) rice maps. GlobCover refers to meshes identified as paddy area by the European Space Agency (ESA) GlobCover land use maps. IRRI+GlobCover refers to meshes identified as paddy area by both IRRI and ESA area maps.
Source: Authors' estimates.

Figure 15: Estimates of Total Planted Area in Ang Thong from Different Data Sources



m^2 = square meter.
Notes: IRRI refers to meshes identified as paddy area by the International Rice Research Institute (IRRI) rice maps. GlobCover refers to meshes identified as paddy area by the European Space Agency (ESA) GlobCover land use maps. IRRI+GlobCover refers to meshes identified as paddy area by both IRRI and ESA area maps.
Source: Authors' estimates.

Figure 16: Estimates of Total Planted Area in Thai Binh from Different Data Sources



m^2 = square meter.
Notes: IRRI refers to meshes identified as paddy area by the International Rice Research Institute (IRRI) rice maps. GlobCover refers to meshes identified as paddy area by the European Space Agency (ESA) GlobCover land use maps. IRRI+GlobCover refers to meshes identified as paddy area by both IRRI and ESA area maps.
Source: Authors' estimates.

It can also be seen that the estimate of total area planted to rice based on the independent measures of rice area using ALIS methodology for all three provinces is higher than the corresponding direct estimates using unmodified and modified track data. However, the mesh level area estimates are closer to official statistics in Ang Thong and Thai Binh, but significantly different from official area estimates in Savannakhet. This can partially be explained by the statistical capacity in the Lao PDR,

since the official yield estimates are also drastically different from crop cutting-derived yield. It is likely that the methodology implemented for estimating rice statistics in the administrative system needs a closer review.[20]

Table 9 shows the weighted estimate of total area in rice based on the mesh-level data on rice area (in m^2) and the corresponding measures of precision. The CV of the independent estimate for Thai Binh is 4.9%, compared to 9.3% for the direct estimate. This indicates an improvement in the level of precision.

The estimate of total area planted to rice in Savannakhet based on the independent measures of rice area in each sample mesh is about 44% higher than the corresponding direct estimate shown in Table 4. The CV of the independent estimate is 7.5%, compared to 16.4% for the direct estimate, so the level of precision is improved considerably. The same is true for Ang Thong where the new estimate of total area planted to rice shown in Table 9 is about 33% higher, and the CV is 4.5%.

These estimates based on the mesh-level rice area data from the digitized Google Earth images could be considered the most accurate estimates of total area planted in rice under the rice crop-cutting pilot survey methodology. These estimates can also be used for improving the level of precision of the estimates of total rice paddy production through ratio estimation, as described in the next section.

Table 9: Estimate of Total Area Planted in Rice Paddy Based on Independent Measure of Total Area Planted with Rice in Each Sample Mesh, Using Digitized Google Earth Images

Domain	Area (m^2)	SE	CV	95% Confidence Interval		DEFF	No. of Observations
				Lower	Upper		
Savannakhet	3,607,317,242	270,231,340	0.075	3,068,869,543	4,145,764,940	1.24	78
IRRI+GlobCover	1,212,448,081	80,751,195	0.067	1,051,547,813	1,373,348,348	0.07	58
IRRI	41,913,872	10,269,821	0.245	21,450,806	62,376,938	0.02	8
GlobCover	1,049,786,538	217,401,053	0.207	616,605,485	1,482,967,592	1.08	10
Other	1,303,168,750	138,331,250	0.106	1,027,537,718	1,578,799,782	0.68	2
Ang Thong	398,383,039	17,855,187	0.045	362,702,295	434,063,784	1.14	66
IRRI+GlobCover	373,212,463	17,069,114	0.046	339,102,562	407,322,365	0.68	50
IRRI	2,639,752	393,107	0.149	1,854,190	3,425,314	0.01	10
GlobCover	22,530,824	5,224,800	0.232	12,089,895	32,971,753	0.21	6
Thai Binh	733,069,437	35,794,877	0.049	661,493,157	804,645,717	1.05	64
IRRI+GlobCover	670,129,990	35,046,872	0.052	600,049,439	740,210,542	0.70	55
IRRI	560,772	252,639	0.451	55,589	1,065,954	0.02	2
GlobCover	62,378,675	7,275,036	0.117	47,831,341	76,926,009	0.07	7

CV = coefficient of variation, DEFF = design effects, m^2 = square meter, SE = sampling error.
Notes: IRRI refers to meshes identified as paddy area by the International Rice Research Institute (IRRI) rice maps. GlobCover refers to meshes identified as paddy area by the European Space Agency (ESA) GlobCover land use maps. IRRI+GlobCover refers to meshes identified as paddy area by both IRRI and ESA area maps.
Source: Authors' estimates.

[20] Similar data consistency issues have also been highlighted in ADB (2017) in the country diagnostic study for the Lao PDR.

F. Improving the Precision of the Estimates of Total Rice Paddy Production through Ratio Estimation

Since we have a more accurate estimate of the total area planted in rice paddy as described in the previous section, it is also possible to obtain a more accurate estimate of the total production of rice paddy by using ratio estimation of the total. This ratio can be expressed as follows:

$$\hat{r} = \frac{\sum_h \sum_i \sum_j W_{3hij} y_{hij}}{6.25m^2 \times \sum_h \sum_i \sum_j W_{3hij}} \tag{12}$$

where:

$\hat{r}$ = ratio estimate of overall rice yield (kg/m^2)

y_{hij} = production in kg of the rice paddy harvested from the crop-cutting data for the sample subplot in the *j*-th sample plot within the *i*-th sample mesh in stratum *h*

The numerator of this ratio is a direct weighted estimate of the total production (in kg) of rice paddy for the province, and the denominator is a direct weighted estimate of the total area planted in rice based on the sample subplots. Since the size of the subplots is uniform (6.25 m^2), this value is simply multiplied by the sum of the weights of all sample subplots. In the case where the rice planted in the sample subplot was completely lost because of flooding or drought, it is possible that the corresponding value of y_{hij} within the sum in the numerator is zero, but the weight of this subplot would still be counted in the denominator.

The direct estimate of the average yield was shown previously in Table 7. The ratio estimate of total rice production would be calculated simply as the independent estimate of the total area planted in rice (from the mesh-level measures of area planted in rice from digitized Google Earth images) multiplied by the overall average yield:

$$\hat{Y}_R = \hat{A} \times \hat{r} \tag{13}$$

In this case, the estimate of the total area in rice ($\hat{A}$) is calculated from the total rice area in each mesh, based on equation 11, and the ratio is the direct estimate of the average rice paddy yield in kg/m^2. On the other hand, the direct estimate of the total rice area in equation 6 is based on the rice area of the individual sample plots and is less precise.

One way to construct a conservative confidence interval for the ratio estimate of the total rice paddy production would be to use the corresponding confidence intervals for the total rice area planted from Table 9 and the average yield from Table 7 and multiply their respective lower and upper limits. This can then be used to estimate the corresponding standard error and CV. The results of the ratio estimation of the total production of rice paddy are shown in Table 10.

Table 10: Ratio Estimate of Total Rice Paddy Production in the Pilot Areas

Domain	Estimate (kg)	SE	CV	95% Confidence Interval	
				Lower	Upper
Savannakhet	693,823,889	93,955,279	0.135	521,700,609	890,005,302
Ang Thong	141,362,508	14,234,430	0.101	114,827,845	170,626,811
Thai Binh	395,857,496	28,887,778	0.073	337,361,510	450,601,602

CV = coefficient of variation, kg = kilogram, SE = sampling error.
Source: Authors' estimates.

The ratio estimate of total rice paddy production for Savannakhet is about 44.5% higher than the corresponding direct estimate of total production in Table 8. Table 10 also shows that the ratio estimate has a lower CV than the corresponding direct estimate of total production of rice paddy. It was also found that the new estimate of total production of rice paddy shown in Table 10 is about 11% lower than the estimate obtained by taking a simple average of rice production in Savannakhet during the rainy seasons of 2010–2014 through official estimates obtained from the CIS (2014).

Meanwhile, the ratio estimate of total rice paddy production for Ang Thong is about 33.2% higher than the corresponding direct estimate of total production. For both Ang Thong and Thai Binh the ratio estimate has a lower CV than the corresponding direct estimates of total production of rice paddy shown in Table 6.

The measures of precision and design effects in the three provinces determined in this study can be useful for determining the sample size that would be needed for nationally representative surveys in each country for measuring the total area and production of rice. In this case, it will be necessary to determine the scope of the survey in terms of the geographic domains to be covered. The sample size will be determined based on a target level of precision for each geographic domain covered by the survey. Here we will describe the procedure for calculating the sample size for future surveys.

The first step would be to determine the possibility of improving the effectiveness of the stratification of the area sampling frame by using more recent and higher-level resolution satellite image data and stratification algorithms, as described in the recommendations. This will help to improve the precision of the survey estimates of total rice area and production for each domain.

The conclusion from the comparison of the alternative area measurement approaches was that the mesh-level measurement of area planted in rice provides the highest level of precision for both the estimates of total area and production of rice. Therefore, the results based on the ALIS methodology can be used for determining the sample size needed to achieve the target level of precision.

For example, a reasonable target for the CV for a key indicator such as the total production of rice in a province is 5%. To determine the required sample size, we can examine the corresponding CV from the pilot survey data based on a sample of 120 meshes. If the sampling methodology for the scaled-up survey will be similar to that of the pilot survey (where four sample plots with rice planted were selected in each sample mesh and one subplot was selected in each sample plot), the design effects from the new survey data would probably not change significantly from those measured using the pilot survey data. In this case, it would be simple to calculate the number of sample meshes that would be required to obtain a 5% CV for a particular indicator from the new survey data. The CV will change based on the square root of the sample size, which in this case can be expressed in terms of the number of sample meshes. This is shown in the following formula:

$$cv_S\left(\hat{Y}\right) = cv_P\left(\hat{Y}\right) \times \sqrt{\frac{n_P}{n_S}}, \tag{14}$$

where:

$cv_S\left(\hat{Y}\right)$ = target coefficient of variation of estimate of total rice production ($\hat{Y}$) for province in the new survey

$cv_P\left(\hat{Y}\right)$ = coefficient of variation of estimate of total rice production for province based on pilot survey data

n_P = number of sample meshes selected for province in the pilot survey (that is, 120)

n_S = number of sample meshes that need to be selected for province in new survey to achieve specified level of precision

Solving this equation to determine the required number of sample meshes, the sample size can be calculated as follows:

$$n_S = n_P \times \frac{\left[cv_P\left(\hat{Y}\right)\right]^2}{\left[cv_S\left(\hat{Y}\right)\right]^2} \tag{15}$$

For example, in the case of Thai Binh Province, the CV for the ratio estimate of the total production of rice from the pilot survey data was 0.073 (7.3%). Using the formula above for calculating the sample size required to obtain a CV of 0.05, we obtain a sample size of 256 meshes. If it is possible to improve the stratification of the sampling frame, the actual CV from the new survey data based on this sample size will probably be lower than the target.

VI. CONCLUSION

Traditional sampling strategies commonly employed in Asia and the Pacific for paddy rice statistics on outdated list frames, incomplete holding information, or administrative data that are prone to numerous biases. The more advanced area sampling method has only been applied in developed economies as it relies on having detailed cadastral information alongside qualified personnel in national statistics offices who are trained in integrating remote sensing techniques and probability sampling methodology. This study explores the use of an area frame multistage stratified sampling methodology to collect paddy rice area and production data in three major rice-producing pilot areas: Savannakhet, Lao PDR; Ang Thong, Thailand; and Thai Binh, Viet Nam, comparing three approaches: (i) a direct estimate obtained through plot measurement using a GPS device, (ii) an alternative direct estimate obtained through digitization of farmer identified plot boundaries on a high-resolution Google Earth image, and (iii) a ratio estimate of total production of rice paddy involving the calculation of the total area planted in paddy rice based on independent mesh-level measures from the digitized Google Earth map. Yield estimates were calculated using crop-cutting techniques. Results from this study suggest that the direct estimates of the total rice paddy area and production from the sample plots have relatively high CVs and wide confidence intervals. The main reason for the relatively high

CVs for the direct estimates of the total area and production of rice paddy is the variability in the size of the sample plots, which are selected within each sample mesh with equal probability. The level of precision of total production of rice paddy can be improved through ratio estimation, after obtaining a more accurate estimate of the total area planted in rice based on the independent mesh-level measures of area planted in rice from the digitized Google Earth maps. This independent measure of total area planted in rice paddy at the sample mesh level reduces the sampling error that results from the variability in sample plot sizes and therefore provides a more precise estimate of total area planted in rice.

We also note that although we used satellite data with the best resolution that was freely available at the time of this study to undergo the stratification process and select random meshes within those strata, we found some inconsistencies in the stratification results. For example, some of the sample meshes visited in the *IRRI + GlobCover* stratum did not have any rice planted. There are two possible explanations for the inconsistencies between satellite-based land cover classification and what was found during the fieldwork: (i) the power of discrimination in the satellite imagery and stratification might not be sufficient or (ii) field teams might not have accurately reported the status of all meshes, thereby systematically excluding some rice-growing meshes from the survey. This indicates that it will be necessary to improve the land use stratification of the frame by using higher resolution satellite images and a greater power of discrimination in the models used for defining the strata. The challenge from the perspective of government agencies is the associated costs of higher resolution data.

Mesh variability is also another important issue. Although the size of the meshes is uniform, the variability in the percentage of area planted in rice inside the meshes increases the CVs of the resulting estimates of total area planted in rice and the corresponding ratio estimates of total rice production. One alternative that can be explored is the possibility of using a different source of satellite data with a stronger discriminatory power for stratifying the meshes. Future work could also test the interviewer effort hypothesis to explain whether a mesh was visited and correctly enumerated by using a logistic regression framework with distance to main road, terrain, slope, enumerator fixed effects, and other covariates as explanatory factors.

The deviation between official statistics for rice area, yield, and production could be due to the presence of nonsampling errors, subjective intervention, and political leadership at the local government levels involving subsequent revisions in the administrative data collection method. Also, yield estimates from crop cutting could be slightly different in Thailand and Viet Nam from official estimates due to the standardization of yield values in our study to 12% moisture content, which is likely to be slightly lower than the moisture content at which farmers sell rice to rice mills or buyers (14%–16%). However, in the Lao PDR, there is almost a doubling of yield estimates from official data compared to crop-cutting results, which warrants further investigation into the existing administrative data collection methods.

Despite these challenges, the use of remote sensing and GIS techniques to obtain rice area and production estimates with relatively high precision is a major reason for the benefit of this methodology compared to the existing administrative data collection system in the Lao PDR, Thailand, and Viet Nam for which measures of precision are not publicly available. With the aid of handheld devices with inbuilt GPS functionalities, the field teams could navigate to the selected meshes, identify plot owners, conduct area measurements, and implement crop cutting. The ESA recently launched the Sentinel-2 satellite which can provide images at 10 m spatial resolution every 5 days for no charge. As satellite data gets cheaper and better, there is a higher likelihood for developing the methodology to adopt area frames.

REFERENCES

Asian Development Bank (ADB). 2016. *Results of the Methodological Studies for Agricultural and Rural Statistics*. Manila.

———. 2017. Lao PDR Accelerating Structural Transformation for Inclusive Growth. Country Diagnostic Study. Manila.

Atzberger, Clement. 2013. Advances in Remote Sensing of Agriculture: Context Description, Existing Operational Monitoring Systems and Major Information Needs. https://doi.org/10.3390/rs5020949.

Beegle, Kathleen, Calogero Carletto, and Kristen Himelein. 2012. "Reliability of Recall in Agricultural Data." *Journal of Development Economics* 98 (1): 31–41. https://econpapers.repec.org/article/eeedeveco/v_3a98_3ay_3a2012_3ai_3a1_3ap_3a34-41.htm.

Biemer, Paul P. 2010. Overview of Design Issues: Total Survey Error. Handbook of Survey Research. ISBN: 978-1-84855-224-1.

Boryan, Claire G., Zhengwei Yang, Patrick Willis, and Liping Di. 2017. "Developing Crop Specific Area Frame Stratifications Based on Geospatial Crop Frequency and Cultivation Data Layers." *Journal of Integrative Agriculture* 16 (2): 312–23. https://doi.org/10.1016/S2095-3119(16)61396-5.

Carfagna, Elisabetta, and Andrea Carfagna. 2010. "Alternative Sampling Frames and Administrative Data: Which is the Best Data Source for Agricultural Statistics?" In *Agricultural Survey Methods*, edited by Roberto Benedetti, Marco Bee, Giuseppe Espa, and Federica Piersimoni, 45–62. West Sussex: John Wiley & Sons Ltd. doi: 10.1002/9780470665480.ch3.

Carfagna, Elisabetta, and Francisco Javier Gallego. 2005. "Using Remote Sensing for Agricultural Statistics." *International Statistical Review 73* (3): 389–404. doi: 10.1111/j.1751-5823.2005.tb00155.

Carletto, Calogero, Sydney Gourlay, and Paul Winters. 2015. "From Guesstimates to GPStimates: Land Area Measurement and Implications for Agricultural Analysis." *Journal of African Economies* 24 (5): 593–628.

Carletto, Calogero, Sara Savastano, and Alberto Zezza. 2013. "Fact or Artifact: The Impact of Measurement Errors on the Farm Size–Productivity Relationship." *Journal of Development Economics* 103: 254–61.

Center for Informatics and Statistics, Ministry of Agriculture and Rural Development (CIS). 2014. *Annual Report*. Ha Noi, Viet Nam.

Cotter, Jim, Carrie Davies, Jack Nealon, and Ray Roberts. 2010. "Area Frame Design for Agricultural Surveys." In *Agricultural Survey Methods*, edited by Roberto Benedetti, Marco Bee, Giuseppe Espa, and Federica Piersimoni, 169–92. West Sussex: John Wiley & Sons Ltd.

Davies, Carrie. 2009. "Area Frame Design for Agricultural Surveys." RDD Research Report Number RDD-09-xx. Washington, DC: United States Department of Agriculture.

De Groote, Hugo, and Oumar Traoré. 2005. "The Cost of Accuracy in Crop Area Estimation." *Agricultural Systems* 84 (1): 21–38.

Deininger, Klaus, Calogero Carletto, Sara Savastano, and James Muwonge. 2011. "Can Diaries Help Improve Agricultural Production Statistics? Evidence from Uganda." *Journal of Development Economics* 98 (1): 42–50.

Faulkenberry, G. David, and Abderrazak Garoui. 1991. "Estimating a Population Total Using an Area Frame." *Journal of the American Statistical Association* 86 (414): 445–49.

Food and Agriculture Organization of the United Nations (FAO). 2012. AQUASTAT Country Profile: Thailand, 2011. http://www.fao.org/nr/water/aquastat/countries_regions/THA.

———. 2002. Rice Information Volume 3 December 2002. http://www.fao.org/docrep/005/Y4347E/y4347e1u.htm.

Fuentes, Montserrat, and Francisco Javier Gallego. 1994. Stratification and cluster estimator on an area frame by squared segments with an aligned sample. *Annual Conference on Applied Statistics in Agriculture* (pp. 112–121). Kansas: New Prairie Press. Retrieved from http://newprairiepress.org/cgi/viewcontent.cgi?article=1353&context=agstatconference.

Gallego, Francisco Javier 2007. "Sampling Efficiency of the EU Point Survey LUCAS 2006." Paper presented at the 56th ISI session, Lisbon.

General Statistics Office (GSO). 2015. *Statistical Yearbook of Vietnam 2014*. Ha Noi: Statistical Publishing House.

Google Earth. http://www.earth.google.com.

Griffin, Richard A. 2014. "Potential Uses of Administrative Records for Triple System Modeling for Estimation of Census Coverage Error in 2020." *Journal of Official Statistics* 30 (1): 177–189. https://doi.org/10.2478/jos-2014-0012.

Grosh, Margaret E., and Juan Munoz. 1996. "A Manual for Planning and Implementing the Living Standards Measurement Study Survey (English)." Living Standards Measurement Study Working Paper 126. http://documents.worldbank.org/curated/en/363321467990016291/A-manual-for-planning-and-implementing-the-living-standards-measurement-study-survey.

Himelein, Kristen, Stephanie Eckman, and Siobhan Murray. 2014. "Sampling Nomads: A New Technique for Remote, Hard-to-Reach, and Mobile Populations." *Journal of Official Statistics* 30 (2): 191–213.

Huddleston, Harold F. 1978. Sampling Techniques for Measuring and Forecasting Crop Yields. U.S. Department of Agriculture. http://ageconsearch.umn.edu/record/142840/files/escs09.pdf.

Kelly, Valerie A., Jane Hopkins, Thomas Reardon, and Eric W. Crawford. 1995. "Improving the Measurement and Analysis of African Agricultural Productivity: Promoting Complementarities between Micro and Macro Data." MSU International Development Paper No. 16. Michigan: Michigan State University. Retrieved from http://ageconsearch.umn.edu/bitstream/54055/2/idp16.pdf.

Khan, Mobushir R., Cornelis A. J. M. De Bie, Herman Van Keulen, Eric Marc Alexander Smaling, and Raimundo Real. 2010. Disaggregating and Mapping Crop Statistics Using Hypertemporal Remote Sensing. https://doi.org/10.1016/j.jag.2009.09.010.

Kilic, Talip, Ismael Yacoubou Djima, and Calogero Carletto. 2017. "Mission Impossible? Exploring the Promise of Multiple Imputation for Predicting Missing GPS-Based Land Area Measures in Household Surveys." Policy Research Working Paper 8138. World Bank. http://documents.worldbank.org/curated/en/668211499349698549/pdf/WPS8138.pdf.

Ministry of Agriculture and Forestry (MAF). 2015. *Agriculture Statistics 2014*. Vientiane.

Sandefur, Justin, and Amanda Glassman. 2014. "The Political Economy of Bad Data: Evidence from African Survey and Administrative Statistics." Center for Global Development Working Paper 373. http://www.cgdev.org/sites/default/files/political-economy-bad-data.pdf.

Singh, Randhir, Ram Chandra Goyal, S. K. Saha, and Raj Chhikara. 1992. "Use of Satellite Spectral data in Crop Yield Estimation Surveys." *International Journal of Remote Sensing* 13 (14): 2583–92.

Singh, Randhir, D. P. Semwal, Anil Rai, and Raj Chhikara. 2002. "Small Area Estimation of Crop Yield Using Remote Sensing Satellite Data." *International Journal of Remote Sensing* 23 (1): 49–56.

Strand, Geir-Harald. 2013. The Norwegian Area Frame Survey of Land Cover and Outfield Land Resources. https://doi.org/10.1080/00291951.2012.760001.Strand, Geir-Harald. 2013. The Norwegian Area Frame Survey of Land Cover and Outfield Land Resources. https://doi.org/10.1080/00291951.2012.760001.

Improving Paddy Rice Statistics Using Area Sampling Frame Technique

This study explores the utility of an area frame developed using remote sensing data in three pilot provinces—Savannakhet in the Lao People's Democratic Republic, Ang Thong in Thailand, and Thai Binh in Viet Nam. It seeks to help address issues related to traditional sampling strategies for paddy rice statistics that rely on outdated list frames, incomplete holding information, or administrative data that may be prone to measurement error. Direct estimates of total paddy rice area and production involving the measurement of plot sizes using Global Positioning System instruments together with a digitally traced map of plot boundaries identified by farmers are also presented in this study, and compared to ratio estimates using independent mesh-level measures.

About the Asian Development Bank

ADB is committed to achieving a prosperous, inclusive, resilient, and sustainable Asia and the Pacific, while sustaining its efforts to eradicate extreme poverty. Established in 1966, it is owned by 67 members—48 from the region. Its main instruments for helping its developing member countries are policy dialogue, loans, equity investments, guarantees, grants, and technical assistance.

ADB

ASIAN DEVELOPMENT BANK
6 ADB Avenue, Mandaluyong City
1550 Metro Manila, Philippines
www.adb.org