Advanced Journal of Agricultural Research and Reviews ISSN: 0410-143X Vol. 2 (1), pp.064-071, February, 2018.
Available online at www.globalacademicsjournals.org © Global Academics Journals

Full Length Research Paper

Feasibility of Dot Sampling Method for Crop Area and Production Estimation: the case of rice in Fogera district , Ethiopia

MinilikTsega[1*], Arouna Aminou[2], Yohannes Fekadu[1], Demeke Negusie[1], DegefieTibebe[1]

[1]EIAR, P.O. Box 2003, Addis Ababa, Ethiopia
[2]Africa Rice, Cotonou, Benin

Accepted 4 January 2017

Lack of reliable and timely information related to crop area and production is one of the major bottlenecks for proper planning and allocation of resources in agricultural development. Generating such information and related datasets in an accurate manner demands statistically sound survey methods. In line with this, the study tries to evaluate the dot sampling method for crop area and production estimation with the objective of looking for simple but sound survey method. Dot sampling tries to estimate planted area by distributing sample dots systematically and directly to a whole target area on the Google Earth and, then, conducting a field survey at the sample dots in the target area. This means the methods does not need sampling frame, interviewing sample households, and conducting objective measurements which needs huge budget and time. In this study, the method was used to estimate rice covered area and the associated yield volume in the Fogera district for 2015/2016 cropping year. The exhibited insignificance difference between the dot sampling estimate and the district reported statistics has indicated the capacity of the dot sampling method to produce reliable crop area and production estimates in relatively short period and minimum cost. Also the advantages and disadvantages of the method were assessed and documented.

Keyword: Dot, Dot sampling, Google Earth, Preparatory survey, Field survey, Estimation

INTRODUCTION

Information related to crop area and production plays important role in planning and allocation of resources for development of agriculture sector. Generating information and related datasets in an accurate manner demands the use of statistically sound survey methods. List and area frame survey methods are the common approaches in use in different countries (Michael Craig and Dale Atkinson, 2013). In Ethiopia, the Central Statistical Agency (CSA) is mandated for production and dissemination of such type of datasets. In this regard, CSA has been collecting agricultural data through annual agricultural sample survey (AASS) for the past three decades using list frame survey method (CSA – ETHIOPIA, 2011).

Usually for the list frame sampling method, census enumeration area (EAs) are the primary sampling units (PSUs). The EA based sampling frame is intended to serve until the next census, which are usually every10 years. A newly completed census cartographic EA list is ideal as sampling frame because it is as current, complete and accurate as any good frame could be. But the longer the interval between the censuses, the less efficient becomes the use of the EAs map (Michael Craig and Dale Atkinson, 2013).This means as time goes, the EA frame itself could be source of biasness unless it is updated regularly. In most cases, EAs will not be updated because of its heavy demand for labor, time and financial resources.

With this survey method, it is also necessary to visit and interview sample households as well to do objective

*Corresponding author Email: miniliktt@gmail.com

measurements of some parameters like planted area and yield. Obviously, both the interview and the objective measurements could be sources of bias. In addition, this survey method usually requires multistage sampling and stratification. For instance, CSA follows a two stage stratified probability proportion to size (PPS) sampling technique, where different administrative zones of the country make the stratum. As the methodology dictates, zone is the lowest reporting level despite massive need for *district* level statistics.

This study presents the evaluation of dot sampling method for crop area and production estimation to alleviate some of the above listed shortcomings of the list frame survey method that is commonly applied in crop area and production estimation.

Dot sampling is a new crop area and production estimation method developed by combining a traditional statistical theory, which is called attribute survey with the latest information technologies, namely Excel and Google Earth (Issei Jinguji, 2014). Unlike to the other conventional methods it does not need sampling frame, interviewing sample households, conducting objective measurements and it can be used at any administrative boundary. It tries to estimate planted area by distributing sample dots systematically and directly to the whole target area on the Google Earth, check the land use attribute at sample points to filter dots, which belongs to the cultivated category, conducting a field survey of the cultivated sample dots in the target area to identify the associated crop land, then estimate the target crop area as a product of total target area and proportion of sample dots associated with the study crop (Issei Jinguji, 2013).
The specific objectives of the pilot study were to produce an estimate of rice area and production in Fogera district in 2014/2015 cropping season, to compare this estimate with figures reported by District and assess the reliability of the dot sampling method as well to explore implementation challenges and benefits and see practicability of the method.

METHODOLOGY

Procedure of the planted area survey using dot sampling method

The entire survey tasks were logically categorized in to four stages namely, study area and sample size determination, preparatory survey, field survey and estimation. Study area and sample size determination stage was focused on delineating actual survey area on the district map and determining the optimum number of sample dots required considering the targeted precision level for the estimates to produce. The preparatory survey stage was mainly concerned on creating sample dots, uploading sample dots on Google Earth, classifying each sample dots into different land use attribute based on visual interpretation of Google Earth image.

The field stage was mainly intended for ground identification of sample dots classified as cultivated in the preparatory survey stage and checking whether the associated crop is rice or not. In this stage of the survey, field practicality of the methodology was assessed considering labor, time, cost and technical demand.
The last stage was designed to estimate the area under rice, which is the product of the total area of the district and the proportion of rice dots. The dot sampling estimate was compared to the district reported statistics to have some clue on the feasibility of the methodology.

Study area delineation and samples size determination

The entire Fogera district was considered for this study. Fogera is located between11º46 to 11º59 latitude North and 37º33 to 37º52 longitude East. It has an area of 1111.4 square kilometer distributed in its 28 kebeles and among which are 19rice producing kebeles. In 7 of rice producing kebeles almost all households produce rice while in the remaining kebeles only few households are rice producers (figure 1). The area under rice in the Fogera district is reported to be 350 square kilometer. The appropriate sample size was estimated to be 3481 dots aimed with 97.5 % precision level for the estimates to be produced (Equation 1).

Where,
n=number of sample dots,P=0.315=proportion of area under rice, Q=1-P=0.685

$$C = \frac{\sqrt{PQ}}{P} : \text{Coefficient of variation in the population}$$

$$d = \sqrt{PQ/n} : \text{Tolerable margin of error}$$

$$CV = \frac{d}{P} : \text{Coefficient of variation in the estimate}$$

$$n = \frac{C^2}{CV^2} = \frac{PQ}{d^2} :\Rightarrow \text{Equation - 1}$$

Sample selection

A rectangular grid, which contains the entire Fogera district was defined on Google Earth using two coordinates which defines the north west and south east coroners of the Fogera map. The north west and south

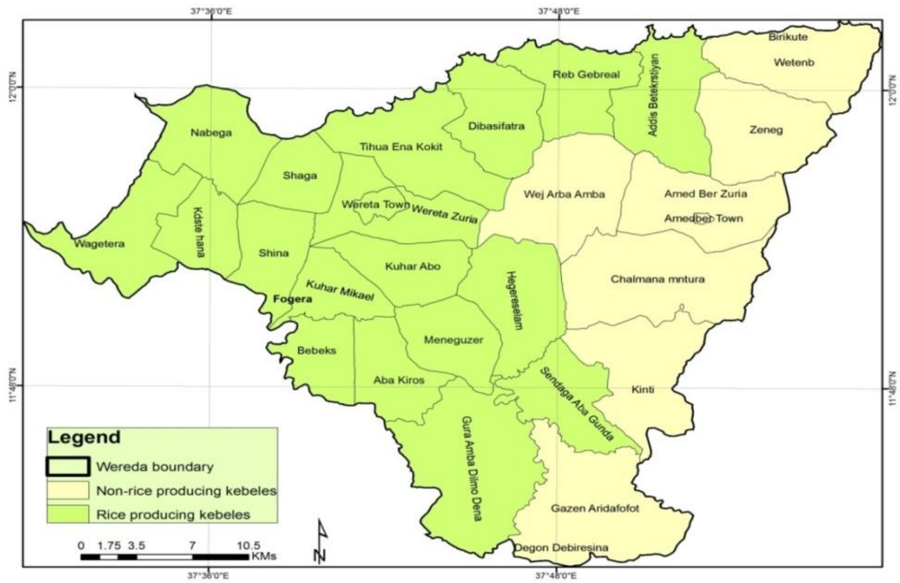

Figure. 1 Study area Fogera district of Amhara region

east corner coordinates of the grid were, respectively used as starting and the ending sample points in a systematic sample selection scheme. The remaining samples were selected by adding a fixed interval on each sample dot starting from the first sample dot. The interval between each sample dot was calculated based on the size of target area and sample size. Since the defined grid also falls outside of the study area (Fogera), the total number of sample points generated were 7032. From these, the total number of sample points that fall within the study area were 3495, which are almost half of the total sample dots generated. The entire sample selection activity was done using a macro program developed for Excel spread sheet (Table 1).

Rice area estimation and comparison

The rational for the estimation lays on the assumption that the proportion of dots in the sample, which fall to given land use category is balanced with the proportion of dots, which belongs to this category in the study population. Based on this assumption the area under rice is calculated as a product of the proportion of rice dots in the sample and area of the district (Equation 2). The estimate obtained was compared with the district reported statistics to have some clue about the reliability of the method. The comparison was done using one sample t-test procedure considering the district statistics as a hypothesized value.

Where, $$\hat{A} = \frac{NRD}{NTD} xTA --- Equation - 2$$

- RD ------------ total number of is rice dots

NTD -------------the total number of dots considered (effective sample size)

TA ---------------land size of Fogera

$\hat{A}$ --------------Estimated rice area

Yield estimation and comparison

For the yield estimation, 36 rice producing households were selected randomly. Of the total rice plots each sample rice producing household cultivate in 2016/2017 cropping season one plot was selected using a simple random sampling technique. From each sample plots two 1mX1m square spots were selected and harvested by enumerators. Selection of the spot-1 was done by moving 30 steps along the edge from a corner of the sample rice plot and enter the plot for 30 steps. For the spot-2, start at the diagonal corner of starting point of spot 1 and find spot 2 using the same procedure as for spot 1. Maximum care was taken to avoid on harvest yield lose and get the actual yield. The average yield was estimated using the following sample mean formula. The estimated yield was compared with the CSA estimates for 2015/16 production season using independent sample t-test procedure.

$$\hat{Y} = \frac{\sum_{i=1}^{32} \sum_{i=1}^{2} y_{ij}}{n} ----Equation - 3$$

where

- $\hat{Y}$ is the yield/m^2 estimate of the district
- y_{ij} is the yield obtained from spot i for the jth sample plot
- n is the total number of spots (64)

Table 1 Dot sampling table combining longitude and latitude coordinates of the sample dots

Name of Longitude→ Latitude↓	37.497832	37.503022	37.508212	37.513402
12.050608	12.050608,37.497832	12.050608,37.503022	12.050608,37.508212	12.050608,37.513402
12.045515	12.045515,37.497832	12.045515,37.503022	12.045515,37.508212	12.045515,37.513402
12.040422	12.040422,37.497832	12.040422,37.503022	12.040422,37.508212	12.040422,37.513402
12.035329	12.035329,37.497832	12.035329,37.503022	12.035329,37.508212	12.035329,37.513402
12.030236	12.030236,37.497832	12.030236,37.503022	12.030236,37.508212	12.030236,37.513402

$$S^2_{(\bar{y})} = \frac{\sum_{i=1}^{n}(y_{ij} - \bar{y})^2}{n-1} ----Equation-4$$

$$CV_{(\bar{y})} = \frac{S_{(\bar{y})} \big/ \sqrt{n}}{\bar{Y}} ----Equation-5$$

Production estimation

Crop production estimates are portrayed as the product of two components area harvested and yield per unit area (Equation 6).

Where,
-----estimated rice production volume
----- estimated yield per hectare
------estimated rice planted area

RESULT AND DISCUSSION

Uploading the sample dots on Google earth

The sample dots were simply uploaded from Microsoft Excel to Google earth using a Microsoft Excel macro program. As the Fogera district administrative boundary map is not a regular rectangular shape, the defined rectangular grid also include areas outside of Fogera. The7032 sample dots distributed regularly over the entire defined grid at a radius of 0.317883 kilometers were uploaded on Google Earth (Figure 2).

Of these 7032 sample dots 3481 fall in the Fogera district, while 14 of them fall on the boundary and the remaining 3537 fall outside of Fogera. Those dots which fall inside and on the boundary of Fogera were considered as the final sample dots of the study (Figure 3).

The sample dots which fall out of the Fogera map were clipped using the Fogera map as a clip feature.

Land use survey on Google earth

The next task after uploading the sample dots was conducting “land use survey” on Google Earth, which is a desk work targeted to identify the land use type at each sample dot location using Google Earth image. The main purpose of the land use survey is to improve efficiency of the field survey by identifying and excluding sample dots which belongs to non-cultivated land use category from the subsequent field survey.

Using Google Earth 2014 image the sample dots were classified as cultivated, non- cultivated, other uses like dyke/river/mountain, tentative reserve and low resolution. In the process, it was possible to identify the land use at 86 percent of the dots. 58, 15, and 13 percent of the dots were classified as cultivated, uncultivated and other use, respectively. This means, at least 58 (64,465ha) percent of the district land was cultivated in 2014/15 cropping season (Table 2). To increase efficiency of the field survey, the identified cultivated dots were again classified as dots belonging to rice and non-rice producing kebeles as it is not possible to get rice dot in non-rice producing kebeles. In total, 1364 cultivated dots which belong to rice producing kebeles were identified.

Field survey

The geographical coordinates of 1364 cultivated sample dots, which belong to the 19 rice producing kebeles were uploaded to GPS. Sketch map showing relative position of the sample cultivated dots in each of the 19 rice producing kebele were produced.

Each of the identified dots were classified as rice, other crop and other land use (Figure 3). In the process 80.6 percent of the dots were found cultivated. This means about 18 percent of the dots were wrongly classified on Google earth as cultivated. Observing this fact, it is also logical to assume some cultivated dots being classified as uncultivated. As backward check shows, these misclassifications were mainly related to the quality of the Google earth image used and on the expertise of the personnel's who did the classification.

Figure 2. Sample dots which fails over the defined rectangular grid

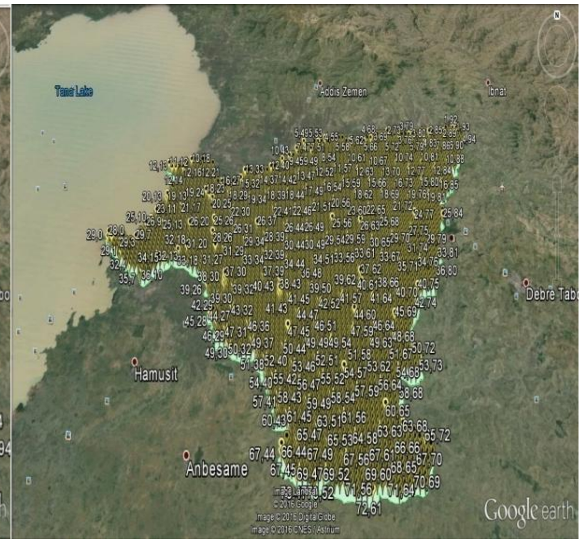

Figure 3.Sample dots which belongs to the study area

Table 2 Result of preparatory survey

Category Name	Category code	Number of Sample	Rate	CV	Area
Non-Cultivated land	1	527	0.15	0.07	16752.42
Cultivated land	2	2037	0.58	0.02	64752.7
Dyke	3	457	0.13	0.07	14527.24
Tentative reserve	4	405	0.12	0.08	12874.25
Low resolution	5	69	0.02	0.20	2193.391
Total		3495			111100

Clearly these misclassifications are sources of both up and downward biases on the preparatory stage estimates. As well on the ground survey, those dots that were wrongly classified as cultivated can be corrected but there is no chance to identify and correct dots wrongly classified as uncultivated. This will always create downward bias on the final estimate. Therefore, it is a must to underline how conditioned the reliability of the dot sampling estimates on the availability of high resolution images. During the ground classification about half of the identified cultivated dots were found to be rice associated. This means rice assumes half of the cultivated area in rice producing kebeles.

Estimation

Rice acreage of 2014/15 cropping season

$$\hat{A} = \frac{\sum_{i=1}^{N} RD_i}{\sum_{i=1}^{N} D_i} xTA = \frac{NRD}{NTD} xTA$$

$$\hat{A} = \frac{550}{3000} x1111.4km^2$$

Coefficient of variation

$$CV = \frac{\sqrt{\frac{\hat{p}\hat{q}}{n}}}{\hat{p}} = \frac{\sqrt{\frac{0.183*0.817}{3000})}}{\frac{550}{3000}} = 0.386$$

Comparison of area estimates

The dot sampling estimate was compared with the district reported statistics using one sample t-test procedure. The test has denied existence of significant difference between the two estimates at 0.05% significance level (Table 3). This result has shown green light to consider the dot sampling method as good candidate for crop acreage estimation. This result can be considered as a call for further detail studies.

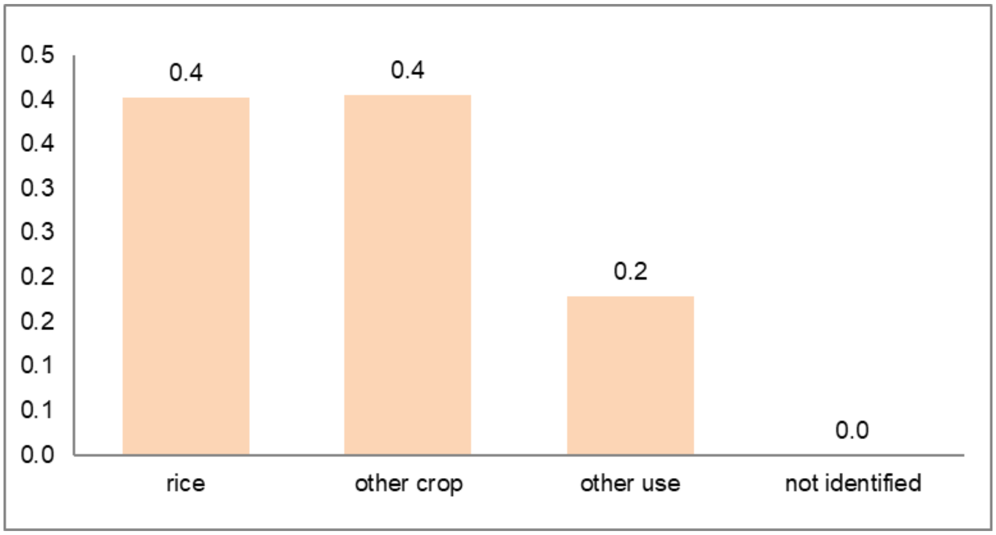

Figure 4. Rice share of the cultivated area in rice producing kebeles

Table 3 Dot sampling estimate against district reported statistics of rice acreage in 2014/2015

Total	Dot sampling	District statistics	t-value	P-value
1111.4	202.74km^2	219.56km^2	-0.444	0.053

Yield Estimation

For this study dry grain yield is defined as grain with 14 % moisture. To this effect, the collected data was adjusted based on the moisture level upon taking weight of the yield obtained from each of the 64 sample spots (Table 4). The conversion factor was obtained from JICA. Using equation 3 and 5 the average dry grain yield and the associated coefficient of variation were estimated to be 3825 kg/ ha and 26.9. $\hat{Y} = \frac{\sum_{i=1}^{32} \sum_{i=1}^{2} y_{ij}}{n} = 3825 kg/ha$

$$CV_{(\bar{y})} = \frac{S_{(\bar{y})} / \sqrt{n}}{\bar{Y}} = 26..9$$

Comparison of the dot sampling and the CSA yield estimates

The reporting level for the CSA estimates is zone. In south Gonder zone there are three rice producing districts one of which being Fogera. Fogera accounts 75% of the rice production in the zone. Having this fact sampling yield estimate is compared with CSA estimate the CSA yield estimate for the zone is considered as indicator value for the Fogera yield level. The dot to have some clue about its reliability. The dot sampling yield estimate is higher than the CSA estimate by 0.8 tons while the CV for the dot sampling is below half of the CSA (Table 5).

Production estimation

The total volume of rice produced in Fogera district for the year 2016/2017 main season was estimated as the product of cultivated area (estimated by dot sampling method) and yield estimated using the crop cutting survey. The 2015/2016 main season rice production was estimated to be 7849 ton.

$$\widehat{P} = \widehat{A} x \widehat{Y} = 20376 ha * 3852 kg/ha$$

$$\hat{P} = 7849 - ton$$

CONCLUSION AND RECOMMENDATION

In this study, the land use survey on Google earth (preparatory

Table 4 Yield obtained from the sample plots

Sample No.	Weight of threshed and clean grains in gm/m2		Moisture content of threshed grains		Weight of dry grains gm/m2		Weight of dry grains kg/ha	
	Spot 1	Spot 2	Spot 1	Spot 2	Spot 1	Spot 2	Spot 1	Spot 2
1	454.6	289.8	14.6	16.4	451	282	4514	2817
2	219.1	374.1	14.9	17.0	217	361	2168	3610
3	290.0	511.3	16.1	15.3	283	504	2829	5036
4	322.4	315.7	15.4	17.4	317	303	3171	3032
5	342.7	279.7	12.2	16.1	350	273	3499	2729
6	533.7	557.2	14.6	27.0	530	473	5300	4730
7	296.4	350.5	14.8	14.8	294	347	2936	3472
8	365.1	476.6	15.0	15.6	361	468	3609	4677
9	237.8	391.8	16.1	17.4	232	376	2320	3763
10	421.8	342.9	20.6	16.2	389	334	3894	3341
11	246.3	288.5	17.0	21.9	238	262	2377	2620
12	335.9	361.3	20.2	22.2	312	327	3117	3269
13	416.0	227.8	25.5	18.7	360	215	3604	2153
14	507.8	589.5	15.1	16.7	501	571	5013	5710
15	459.4	555.0	19.5	15.9	430	543	4300	5427
16	564.4	316.1	17.4	15.4	542	311	5421	3109
17	421.6	659.6	15.7	15.7	413	647	4133	6465
18	299.7	365.8	26.6	24.6	256	321	2558	3207
19	464.1	553.3	16.5	12.8	451	561	4506	5610
20	301.9	377.1	17.5	17.1	289	364	2886	3635
21	343.1	329.8	13.2	16.2	346	321	3463	3214
22	619.0	684.0	19.4	20.4	580	633	5801	6331
23	442.1	454.4	18.8	17.7	419	435	4195	4349
24	457.1	420.4	18.9	15.4	431	414	4310	4135
25	477.6	425.4	18.2	18.2	454	405	4543	4046
26	439.7	429.1	14.2	15.9	439	420	4387	4196
27	374.4	332.4	14.0	19.7	374	310	3744	3104
28	305.2	266.2	12.7	15.4	310	262	3098	2619
29	410.5	574.9	15.8	17.1	402	554	4019	5542
30	554.4	481.5	18.0	18.2	529	458	5286	4580
31	321.4	287.5	15.4	19.7	316	268	3162	2684
32	406.4	399.2	16.3	16.8	396	386	3955	3862
33	362.2	375.1	13.9	13.5	363	377	3626	3773
34	267.2	312.4	11.5	11.8	275	320	2750	3204
35	421.6	317.2	13.1	12.8	426	322	4260	3216
36	267.1	290.3	16.4	18.2	260	276	2596	2761

survey) has exhibited its potential to estimate reasonably the different land use status of a given area if quality and up to date image is used. This means, it can be used to validate district reported statistics and estimates from

Table 5 Comparing the dot sampling yield estimate against the CSA estimates (ton)

Dot sampling		CSA	
Yield	CV	Yield	CV
3.8	26.9	3.1	61.0

other methods as long as high resolution image is available.

The field survey was a bit challenging in some of the sites because the topography of the area where some of the dots fall, unavailability of road access to some of the dots, and long distance walking demand. Beside the mentioned short comings, discrepancy between on ground and Goggle Earth land use classification was evidenced. As backward check shows, some of these misclassifications were mainly related to the quality of the Google earth image used and on the expertise of the personnel's who did the classification. Therefore, for a better result the time the image was taken, quality of the image to be used and the enumerators image interpretation capacity should be considered seriously.

The exhibited insignificance difference between the dot sampling estimate and the district reported statistics has indicated the capacity of the dot sampling method to produce reliable crop area estimates in relatively short period and minimum cost. Based on this result we need to do further detail study and see the possibility to consider dot sampling as alternative method for crop acreage estimation.

The dots sampling method can be used to produce local level crop area estimates as this study did for Fogera which is below AASS reporting level. It can also be used jointly with yield per unit area estimates from other methods to produce reliable production estimates at any administrative level. In general, it can be concluded that dot sampling method has the potential to be a quick fix to produce local level production and area statistics, which is not available from national level surveys like AASS.

Usually, in national level surveys as AASS rare crops, like rice in Ethiopia, are not well represented and consequently precision level of the estimates associated with this type of crops is low. In this regard, dot sampling method can be a better option to conduct a crop level standalone surveys to produce area and production estimates with the required precision level.

REFERENCE

CSA – Ethiopia (2011). Comparative Study Between Area Frame and List Frame for Agricultural Survey, For African Commission on Agricultural Statistics Twenty-second Session

Issei Jinguji (2013).Rice Production Survey using the Dot Sampling Method and Google Earth, Improving Food Security Information in Africa. Africa Rice Center, Ministry of Agriculture, Forestry and Fisheries, Japan.(July 2013),

Issei Jinguji (2014). Kamakura city, Japan Expert meeting on ‘Crop Monitoring for Improved Food Security’ in Vientiane. FAO Regional Office for Asia and the Pacific (FAO RAP) 17 February 2014, Vientiane, Lao PDR

Michael Craig and Dale Atkinson (2013). A Literature Review of Crop Area Estimation, July 2013, For UN-FAO.