Technical Report

Improving methods for estimation of crop area, yield and production under mixed cropping

Under the Project:
Research on Improving Methods for Estimating Crop Area, Yield and Production under Mixed, Repeated and Continuous Cropping

Funded by:
Global Strategy to Improve Agricultural and Rural Statistics

Importance of Mixed and Continuous Cropping

- **The practice of sowing crops in mixture in same parcel of land is prevalent in almost all countries, particularly where land holdings are small.**
- **It provides protection to the farmers in adverse weather conditions *i.e.* drought, flood, pest and disease infestation, enables utilization of available space, moisture and nutrients in the field.**
- **Cultivators usually mix crops that cannot stand a particular type of weather with another set of crops that resist and thrive under the same conditions.**

Contd…

- Generally, when crops are grown mixed, recording of area under each component crop at field level becomes difficult.
- Ignoring crops grown mixed i.e. accounting only the principal crop may lead to underestimated production
- The crop mixture may vary field-wise and even seed ratio of component crop may also vary in each crop mixture resulting in high variability.
- Due to high variability in the data, estimation of crop area and yield under mixed require fairly dispersed samples for precise estimation
- The use of only objective methods for crop area and yield estimation may create implementation problems

Importance of Mixed and Continuous Cropping

- **Sowing and harvesting same or different crops on the same piece of land one after the other during the agricultural year is called as *continuous cropping*.**
- **Provides a way to grow more than one crop on the same piece of land.**
- **More common among the small holders as it provides repeated use of land.**
- **Area planted gradually increased-serious problem**
- **Likely to complicate the problem of estimation of crop area and yield**

Contd…

- **Appropriate methodology needed for estimation of crop area and yield under mixed and continuous cropping.**
- **Study entitled "Improving Methods for Estimating Crop Area, Yield and Production under Mixed and Continuous Cropping" awarded to the ICAR-Indian Agricultural Statistics Research Institute (ICAR-IASRI) under the Global Strategy to Improve Agricultural and Rural Statistics.**

Objectives of the Study

- **To critically review the literature pertaining to crop area and yield under mixed and continuous cropping (TR1).**
- **To identify the gaps relating to estimation of crop area and yield under mixed and continuous cropping.**
- **To develop a standard statistical methodology for estimation of area and yield rate under mixed and continuous cropping (TR2).**
- **To test the developed methodology in three field testing countries one each in Asia-Pacific, Africa and Latin America/Caribbean region. (FTP, TR4)**
- **To identify issues and challenges and provide suitable guidelines for implementation of the developed methodology in developing countries.**

Approach Used in the Study

- **Combined subjective and objective methods for crop area and yield estimation used**
- **Sample survey approach used**
- **Methodology developed in three scenarios**
  - **Cadastral map approach**
  - **Area frame approach**
  - **Household approach**

Testing of Developed Methodology

- **The developed methodology was field tested in one district/study area each in Indonesia, Rwanda and Jamaica**
- **Household approach used in Indonesia and Jamaica**
- **Area frame approach used in Rwanda**

Name of identified countries	Province	Name of identified districts/study area
Indonesia	West Java	Cianjur
	Yogyakarta DI	Gunungkidul
Rwanda	Northern	Musanze
Jamaica	Two parish: Trelawny Manchester	Study Area (consisted of 5 Extension Areas i.e. Albert Town, Christiana, Comfort Hal, Lowe River, Warsop

Table-1.3: Name of countries along with country wise name of identified districts / study area

Data Collection Approach

- **PAPI method used in all three countries**
- **CAPI method used in Indonesia and Jamaica**
- **CAPI method used in Gunungkidul district for listing, crop area and yield estimation**
- **CAPI method used for listing and crop area estimation in Jamaica. PAPI method used for crop yield estimation**
- **PAPI method used in listing, crop area and yield estimation in Rwanda**

Measurement Methods Used

- **Crop Area Measurement**
  - **Auxiliary variables: Farmer self reported area, Total number of family members of household, Number of active family members, Map area**
  - **Study variables: GPS, Enumerator reported area**
- **Crop Yield Measurement**
  - **Auxiliary variables: Farmer predicted produce, Farmer recall of produce**
- **Study variables: CCE, sampling of harvest units, Expert assessment of crop produce, Enumerator reported produce**

Results of field testing in three countries

Indonesia

- **The Badan Pusat Statistik (BPS), Indonesia is the designated agency for collection of agricultural statistics.**
- **It is responsible for conducting major censuses and surveys as also for dissemination of all official statistics in the country.**
- **In view of its expertise in the area of agricultural statistics, BPS, Indonesia was identified as the nodal agency for data collection work under this study**

Mixed Crop Situation in Gunungkidul District

Main Crops

- **Dryland paddy, Wetland paddy, Maize, Soybean, Cassava and Peanut**

Major Crop Mixtures

- **Dryland paddy with Maize, Cassava, Soybean**
- **Dryland paddy with Soybean and Peanut**

A crop in a mixture was excluded if it occupied less than 10% area in the mixture, crop area was allocated to dominant crop in the mixture.

Crops Targeted in Indonesia & Sampling Frame Used

- **Targeted crops: Wetland paddy, Dryland paddy, Maize, Soybean. Mixture of these crops and the four crops mixed with other crops**
- **Sampling frame: List of eligible households formed the sampling frame**
- **Eligible household: at least one parcel growing the study crop or mixture of these crops or these crops mixed with other crops**
- **All eligible parcels of eligible households were considered for crop area measurement**

Sampling Design Used for Sample Selection for Crop Area Estimation

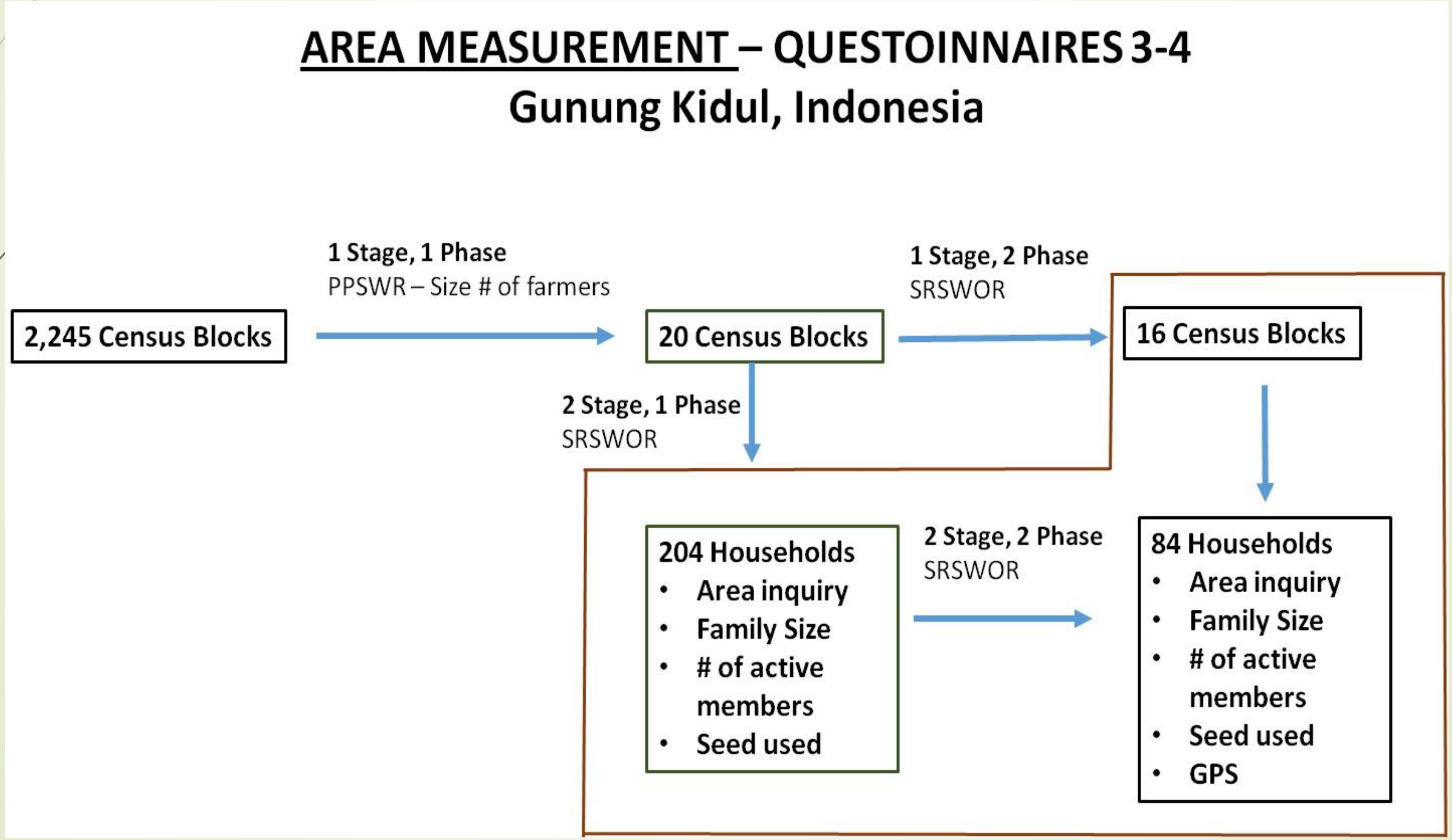


Table 2.1.5.1: District wise, Stratum wise Number of Selected Census Blocks & number of Selected Households in Selected Census Blocks in different phases under Study

Name of Province	District	Strata	Total Number of CBs	Number of eligible HHs in selected CBs for Inquiry	Number of Selected CBs / HHs/parcels							
					CBs				HHs/parcels			
					Area		Yield		Area (HHs)		Yield (Parcel)	
					Inquiry	GPS	Assessment	CCE etc.	Inquiry	GPS	Assessment	CCE
DI Yogyakarta	Gunung kidul	1	289	321	4	3	3	2	40	15	21	12
		2	449	382	5	4	4	3	52	22	28	16
		3	643	413	5	4	3	2	51	20	24	11
		4	864	420	6	5	5	4	61	27	28	16
Total			2,245	1536	20	16	15	11	204	84	101	13
West Java	Cianjur*	1	3,501									
		2	281									
		3	665	77	1	1	#	#	40	40	#	#
		4	2036									
Total			6,483	77	-	1			40	40		

Data on plant count could not be collected

Period & Method of Data Collection

Period of Data Collection

- The data on crop area, yield and production was collected from Gunungkidul: December, 2015 to April 2016
- Cianjur district: data collected from December, 2015 to March 2016.
- Crop failure led to data not being collected in Cianjur district on crop yield

Method of Data Collection

- Cianjur: PAPI method
- Gunungkidul: CAPI method
- Questionnaires and Instruction manual prepared accordingly

Results & discussions - Indonesia

- **Crop area estimation**
  - **Variation in parcel size from 0.01 ha to 1.31 ha by GPS while by**
  - **Farmers' inquiry variation is from 0.01 ha to 1.50 ha**
  - **% rel. difference between farmers' inc and GPS area decreases as the parcel size increases.**

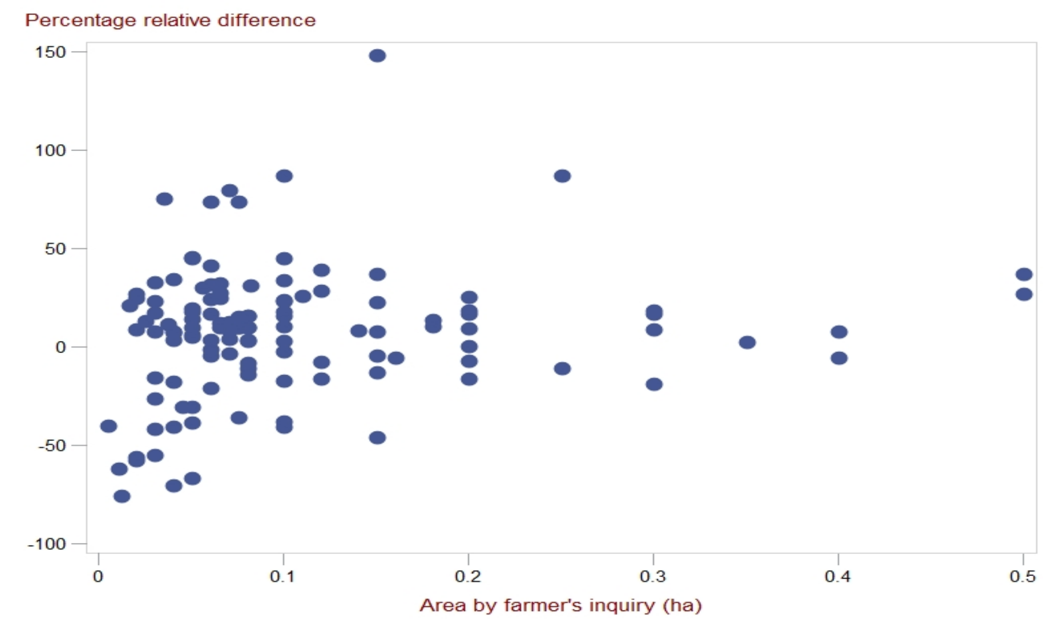


Figure: 2.1.8.1 Percentage relative difference between the average GPS measure and area given by farmer inquiry

Results & Discussions – Indonesia

- **Correlation between all the four variables is very high**
- **Maximum correlation is found between crop area by GPS & Inquiry**

1.00	0.96	0.86	0.85
0.96	1.00	0.90	0.87
0.86	0.90	1.00	0.96
0.85	0.87	0.96	1.00

Table 2.1.8.1: Correlation matrix indicating correlation between Crop area by GPS, Crop area by Inquiry, Active family members and total number of family members.

Note: sampling weights incorporated in computing correlation coefficients.

Results & Discussions – Indonesia

- **GPS readings as study variable and farmers' inquiry based crop area as an auxiliary variable is most precise estimator.**
- **Apportioning of crop mixture area into component crops was achieved using the information on seed rates. Farmers could recall the seed use for each parcel but not for each subparcel**
- **Small sample sizes explain the high percentage standard errors.**

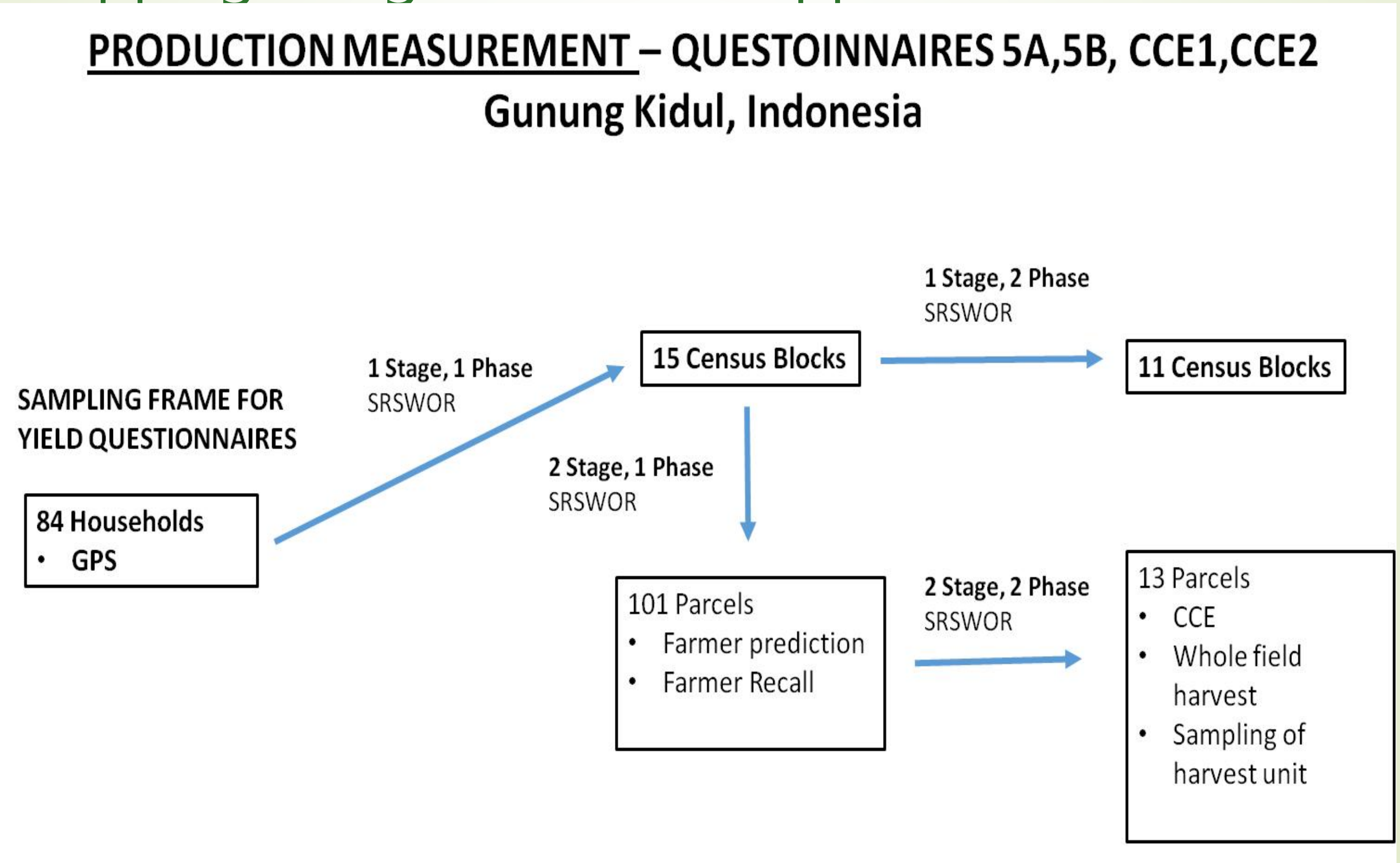
Crop name	Crop wise sample sizes	Crop Area Estimates By GPS only		Crop Area Estimates using total number of family members as auxiliary variables along with % SE		Crop Area Estimates using active family members as auxiliary variables along with % SE		Crop Area Estimates using Crop Area by inquiry as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

Table 2.1.8.2: Crop wise regression estimator of Crop Area by GPS using Crop Area by Inquiry, Active family members and total number of family members as auxiliary variables.

Crop Yield Estimation

- **The sample selected for crop area estimation provided the sampling frame for selection of sample for crop yield estimation**
- **A stratified two-stage sampling design with two phases at each stage of sampling was used for the purpose of selection of the sample for crop yield estimation**
- **PSUs: Census blocks**
- **SSUs: Parcels growing the study crop/various study crops mixed together/study crops mixed with other crops**
- **Mixture by mixture sampling at second stage**

Sampling Design for Crop Yield Estimation for Mixed Cropping using Household Approach



Crop Yield Estimation – Comparison of Various Methods Against Gold Standard

- **Sampling of harvest unit method consistently under-estimated the production**
- **CCE method consistently over-estimated the production**
- **Between the two, the magnitude of deviations were more in CCE method**

Crop name	Number of parcels	Percentage relative difference between CCE and whole field harvest	Percentage relative difference between sampling of harvest unit and whole field harvest
Dryland Paddy	10	178.083	–4.520
Wetland Paddy	2	46.122	–28.689

Table 2.1.8.3: Percentage relative differences between methods of Crop Cutting Experiment, Sampling of Harvest Unit from Whole Field Harvest

Note: Comparisons based on small sample sizes.

Estimates of Crop Yield along with Percentage Standard Errors

- **Recommendations: the method of sampling of harvest unit along with farmer prediction of crop produce as auxiliary variable is most precise estimator.**
- **Recommendation criteria: Closeness to gold standard and lower percentage standard errors.**

Estimators	Auxiliary variable	Estimates (Kg/ha)	%SE
Crop yield by CCE	Farmer prediction of crop produce	3250	25.84
Crop yield by CCE	Farmer crop recall of produce	2390	34.96
Crop yield by sampling of harvest units	Farmer prediction of crop produce	2087	24.14
Crop yield by sampling of harvest units	Farmer crop recall of produce	1327	38.36

Table 2.1.8.4: Estimates of crop yield along with percentage standard error using double sampling regression estimators involving different variables.

Note: Estimates based on small sample sizes.

Optimum Value of Sample Size

Cheaper the cost of sampling the auxiliary variable vis-à-vis the main variable, greater the cost reduction in the double sampling regression estimator over the linear estimator.

$C_2/C_1=10$			Percentage reduction in cost	$C_2/C_1=15$			Percentage reduction in cost
n	n'	n_1		n	n'	n_1	
143	153	175	7.21	125	132	173	20.14

Table 2.1.8.5: Number of census blocks to be selected for farmer predicted dry land paddy yield and for sampling of harvest unit under system II over that of system I respectively and percentage reduction in cost in adopting of system II over that of system I for different values of cost ratios in the Gunung Kidul district for dry land paddy (5%).

Minimizing the cost keeping the % SE fixed

Note: Recommendations based on small sample sizes.

Results & Discussions - Rwanda

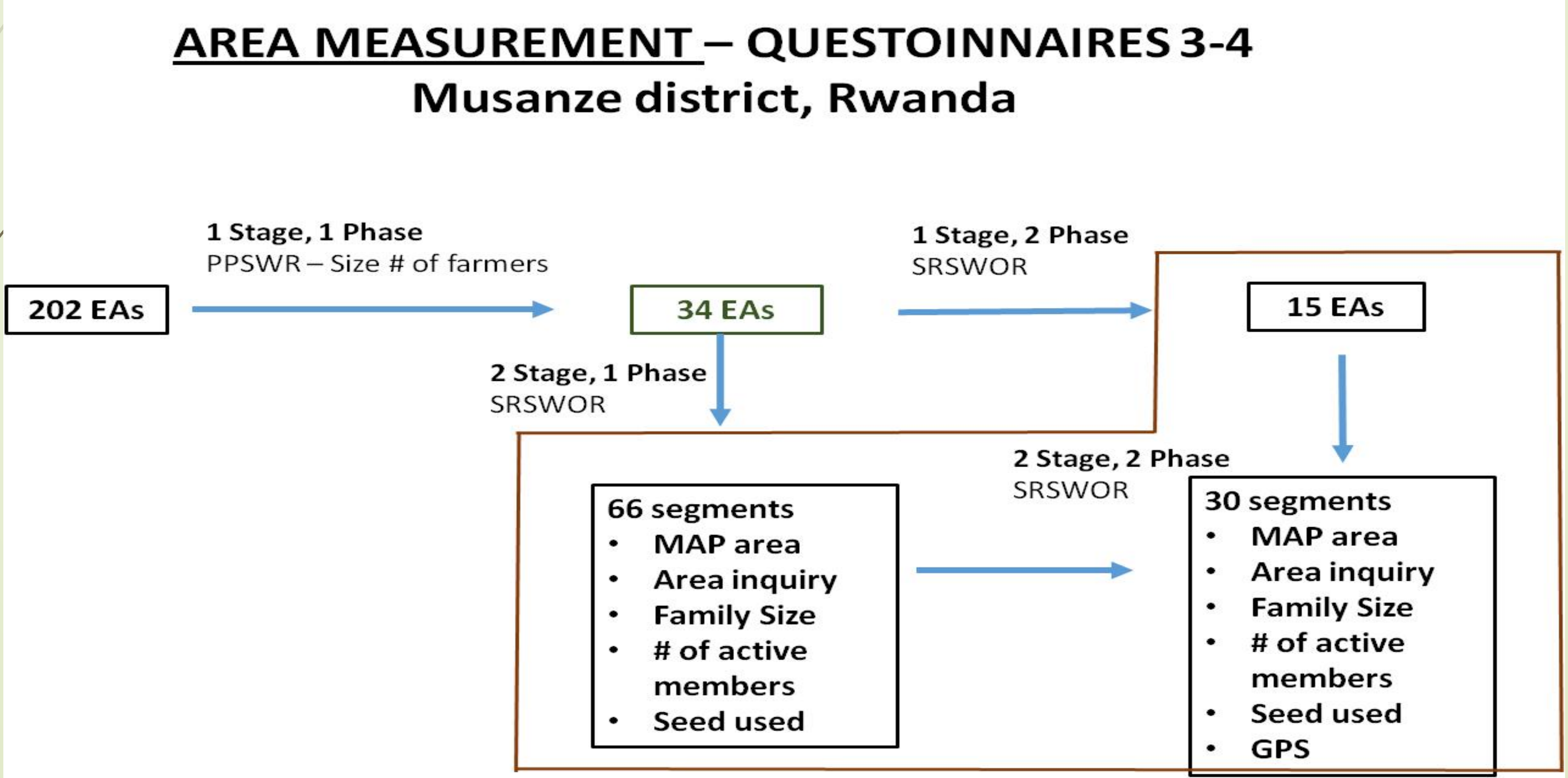
Mixed and continuous cropping situation in the country

- **National Institute of Statistics Rwanda (NISR), Rwanda identified Musanze district for field testing.**
- **The main food crops grown in the district are beans, maize, Irish potato and sorghum.**
- **Common crop mixtures in the district are maize with beans.**

Sampling Frame Used

- **The district divided into 2 strata**
- **Stratum I: 196 EAs Stratum II: 6 EAs**
- **Size of an EA: 100 ha each divided into 20 segments of 5 ha each**
- **Segments are divided into tracts owned and operated by farmers**
- **16 EAs selected from Stratum I and 4 from Stratum II by PPSWR**
- **Size measure used was agricultural area**
- **2 segments selected from each selected EA using SRSWOR design and were completely enumerated to identify eligible tracts**

Sampling design for crop area estimation for mixed cropping when Area Frame is available



Description of Samples for Data Collection

Strata	Total Number of EAs	Total Number of selected EAs for area by map	Total Number of segments in 34 selected EAs	Number of selected segments in 34 selected EAs for area by map	Number of eligible tracts in selected segments for area by map	Number of Selected EAs / tracts/parcels						
						EAs (Segments)			tracts/parcels			
						Area	Yield		Area (tracts)		Yield (parcel)	
						Inquiry and GPS	Assessment	CCE etc.	By MAP	GPS	Assessment	CCE
1	196	30	1061	59	2349	12(24)	11	8	814	349	179	25
2	6	4	132	7	403	3(5)	3	2	143	90	21	7
Total	202	34	1193	66	2752	15(29)	14 (27)	10	957	439	200	32

Table 2.2.5.1: Stratum wise total number of EAs and segments, Number of Selected EAs and selected segments in selected EA and number of eligible and selected tracts within selected segments and Selected EAs in different phases

Period & Methods of Data Collection

Period of Data Collection

- Data collection period: January to March, 2016.
- Listing exercise completed in January 2016.
- Data collection on the different variables carried out during February to March, 2016.

Method of Data Collection

- PAPI method used for data collection
- NISR, Kigali, Rwanda was the data collection agency

Crop Area Estimation

The GPS measurements range from 0.0004 ha to 0.3522 ha. The range of variation in ESR was 0.0012 ha to 0.4 ha. Parcel area in Rwanda smaller than Indonesia.

The percentage relative difference between ESR & GPS decreases as parcel size by ESR increases

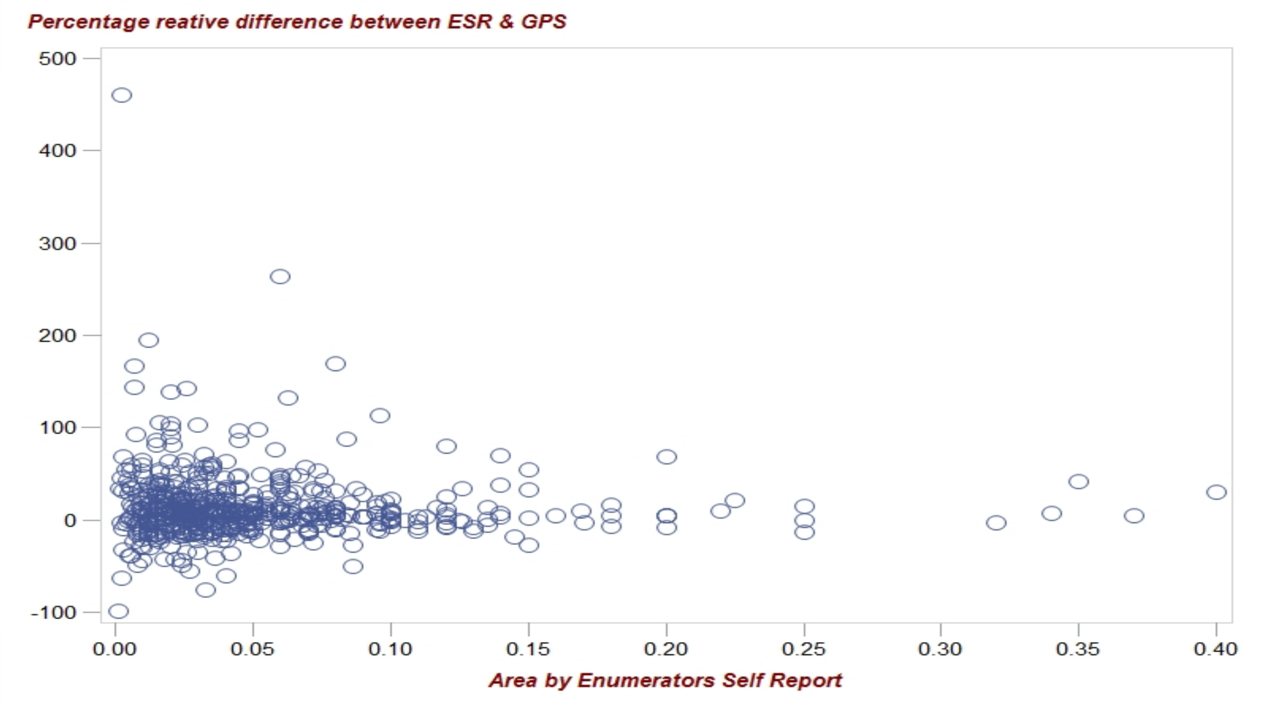


Figure 2.5.2.8.1: Percentage relative difference between the Enumerators Self-Reported area and GPS in Musanze district.

Crop Area Estimation

Percent relative difference between MAP & GPS decreases as the parcel size through MAP increases.

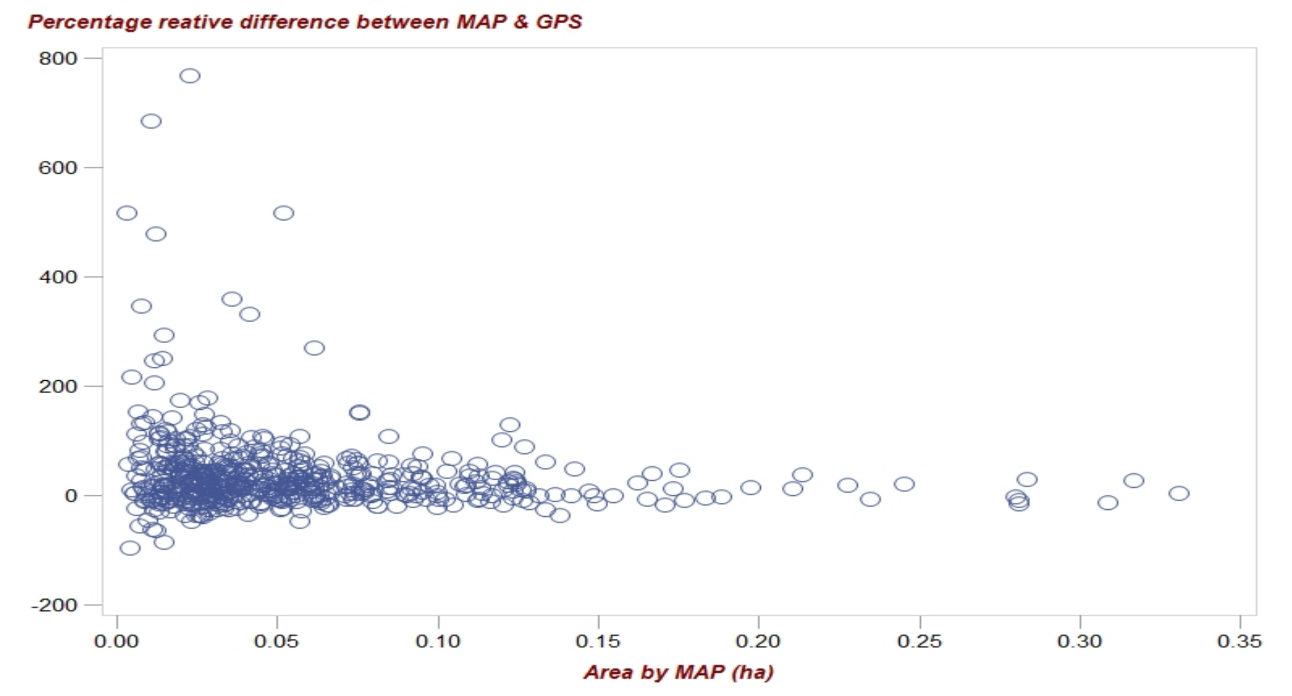


Figure 2.5.2.8.2: Percentage relative difference between the average GPS measure and area by Map in Musanze district.

Crop Area Estimation

Area_class_gps	Number	Avg_diff_Map_GPS	Avg_rel_diff_Map_GPS	Avg_diff_ESR_GPS	Avg_rel_diff_ESR_GPS
1	518	0.007865	**40.6004**	0.003171	**15.3737**
2	46	0.000338	**-0.1167**	-0.00202	**-1.2227**
3	7	0.005795	**2.484**	0.006076	**2.4928**
4	4	-0.02582	**-7.746**	0.031412	**10.0683**
All	575	0.007003	**36.5427**	0.002987	**13.8523**

Table 2.2.8.3: Area class wise average relative difference between parcel area as measured by GPS vis-à-vis area by Map and Enumerator self reported area

Average relative difference between Map area and GPS area as well as enumerator self reported area and GPS area declines as the parcel size increases. Enumerator self reported parcel area is closer to the GPS area compared to the map based area.

Note: Area class: 1: <= 0.1 ha, 2: >0.1 & <=0.2, 3: >0.2 & <= 0.3, 4: >0.3

Correlation matrix between different variables

The GPS, Map and Enumerator self reported crop areas are highly correlated.
The variables number of family members and active family members also exhibit good correlation among themselves as well as with GPS, Map and enumerator self reported area.

	MAP	GPS	Enumerators Self Reporting	Family member	Active family member
MAP	1.00	0.99	0.99	0.70	0.69
GPS	0.99	1.00	0.99	0.78	0.77
ESR	0.99	0.99	1.00	0.74	0.73
FAM	0.70	0.78	0.74	1.00	0.99
ACT	0.69	0.77	0.73	0.99	1.00

Table 2.2.8.4: Correlation matrix indicating correlation between Crop area by Map, crop area by GPS, Enumerator reported crop area, number of family members and active family members

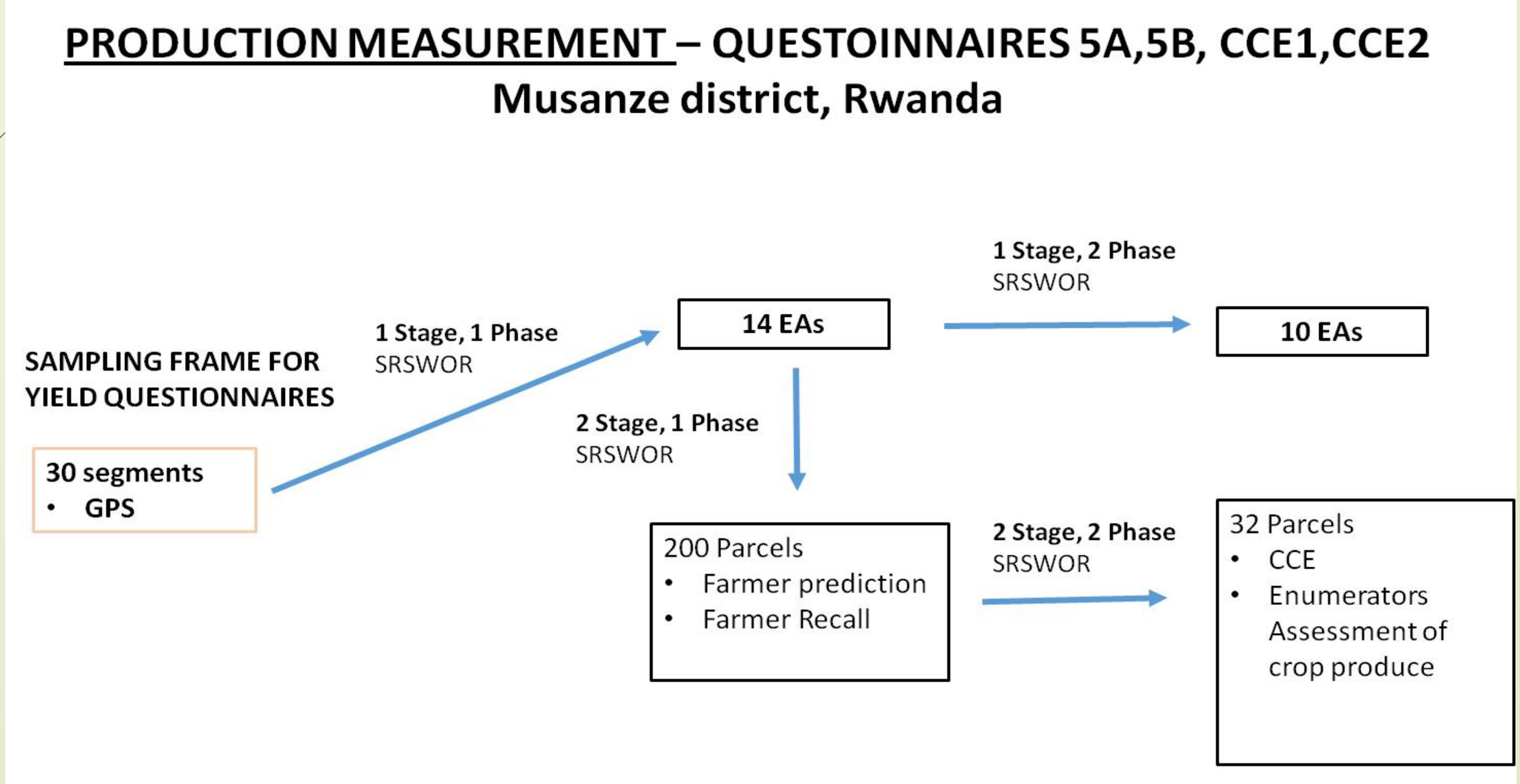
Note: sampling weights incorporated in computing correlation coefficients.

The regression estimator involving ESR based crop area as the dependent variable and Map based crop area as auxiliary variable is the most reliable. The regression estimator in which GPS crop area is used as the dependent variable and Map based crop area as independent variable is more or less equally precise and is recommended for use.

	Crop Area Estimates		Crop Area Estimates		Crop Area Estimates		Crop Area regression Estimates		Crop Area regression Estimates		Crop Area regression Estimates		Crop Area regression Estimates		Crop Area regression Estimates		Crop Area regression Estimates	
	By GPS only		By Enumerators Self reported area only		By Map only		using Map as auxiliary variables along with % SE		using Map as auxiliary variables along with % SE		using Total family member as auxiliary variables along with % SE		using Total family member as auxiliary variables along with % SE		using Active member as auxiliary variables along with % SE		using Active member as auxiliary variables along with % SE	
	Crop Area Estimates (ha)	%SE	Crop Area Estimates (ha)	%SE	Crop Area Estimates (ha)	%SE	Crop Area Estimates (GPS) (ha)	% SE	Crop Area Estimates (ESR) (ha)	% SE	Crop Area Estimates (GPS) (ha)	% SE	Crop Area Estimates (ESR) (ha)	% SE	Crop Area Estimates (GPS) (ha)	% SE	Crop Area Estimates (ESR) (ha)	% SE
CLIMBING BEANS	4790.81	22.59	5251.59	22.34	5560.33	22.35	4125.51	17.20	4504.06	17.19	4173.76	25.76	4638.61	25.14	4302.52	24.98	4776.27	24.41
MAIZE	13169.02	24.70	14640.78	24.16	15652.88	23.64	11182.26	23.83	12508.98	23.31	11954.39	27.15	13402.66	26.34	12019.59	27.00	13465.38	26.21
SWEET POTATOES	168.95	63.32	186.73	59.84	207.35	55.16	144.33	47.59	160.92	44.38	148.17	71.95	164.61	67.64	144.51	73.76	160.67	69.29

Table 2.8.2.5: Double sampling approach based estimators as well as simple linear estimators of crop area along with their percentage standard errors

Sampling design for crop yield estimation for mixed cropping when Area Frame is available



Comparison of crop yield estimation methods

- **Between the CCE and enumerators assessment of crop produce methods, the enumerators' assessment of crop produce method is closer to the whole field harvest.**

Q_2_5_1_Crop_mixture_name	_TYPE_	_FREQ_	avg_Rel_diff_CCE_WFH	avg_Rel_diff_EAP_WFH
	1	32	-13.107	-8.4813
Climbing beans	1	8	35.946	34.3611
Maize	1	21	186.267	-6.7739

Table 2.2.8.6**:** Percentage relative differences between method of Crop Cutting Experiment, Enumerators assessment of produce from Whole Field Harvest

Correlation matrix between different variables

The variable Enumerator assessment of crop produce is highly correlated with farmer predicted crop yield and crop yield through farmer recall method.
Crop cutting experiment exhibits moderately high correlation with farmer predicted crop yield and crop yield through farmer recall.
High correlation is observed between farmer predicted crop yield and crop yield through farmer recall.

variable	yield_CCE	yield_EAP	yield_5A	yield_5B
yield_CCE	1.00	0.72	0.59	0.42
yield_EAP	0.72	1.00	0.90	0.87
yield_5A	0.59	0.90	1.00	0.89
yield_5B	0.42	0.87	0.89	1.00

Table 2.2.8.7: Correlation Matrix indicating correlation between Enumerators Assessment of crop produce, Crop cutting experiment produce, farmer prediction and farmer recall of produce

Double Sampling Approach Based Regression estimators

Crop_mixture	Crop_name	Crop yield by CCE		Crop yield by Enumerators Assessment of produce		Crop yield by CCE				Crop yield by Enumerators Assessment of produce			
						Farmer prediction of crop produce		Farmer recall of crop produce		Farmer prediction of crop produce		Farmer recall of crop produce	
		Est. (kg/ha)	%SE	Est. (kg/ha)	%SE	Est. (kg/ha)	%SE	Est. (kg/ha)	%SE	Est. (kg/ha)	%SE	Est. (kg/ha)	%SE
Climbing beans	Climbing beans	338.2	60.3	565.1	24.2	425.3	40.7	513.2	33.7	623.6	18.6	682.6	16.9
Maize	Maize	2181	19.6	812.8	21.0	2014.2	20.3	2061.6	20.7	691.5	21.1	580.0	28.2
Climbing Beans+ maize	Maize	232.5	70.6	308.7	89.1	258.3	55.7	286.5	49.3	353.5	67.4	403.7	57.5

Table 2.2.8.8: Estimates of crop yield along with percentage standard error using simple estimators and double sampling regression estimators involving different variables.

Contd…

- **The double sampling approach based regression estimator is recommended here when Enumerator assessment of crop produce is taken as main variable while crop produce through farmer recall method is taken as auxiliary variable.**
- **The not so good performance of double sampling regression estimator vis-à-vis the linear estimator in maize crop is attributable to small sample sizes used coupled with high variability in the data.**

Recommended Optimum Sample Sizes

Climbing Beans											
$c_{2/}c_1$=5			Percentage reduction in cost	$c_{2/}c_1$=10			Percentage reduction in cost	$c_{2/}c_1$=15			Percentage reduction in cost
n	n'	n_1		n	n'	n_1		n	n'	n_1	
23	69	92	66.67	20	78	92	81.99	10	74	92	87.29
Maize											
824	590	837	-	820	635	837	-	814	670	837	-

Results & Discussions - Jamaica

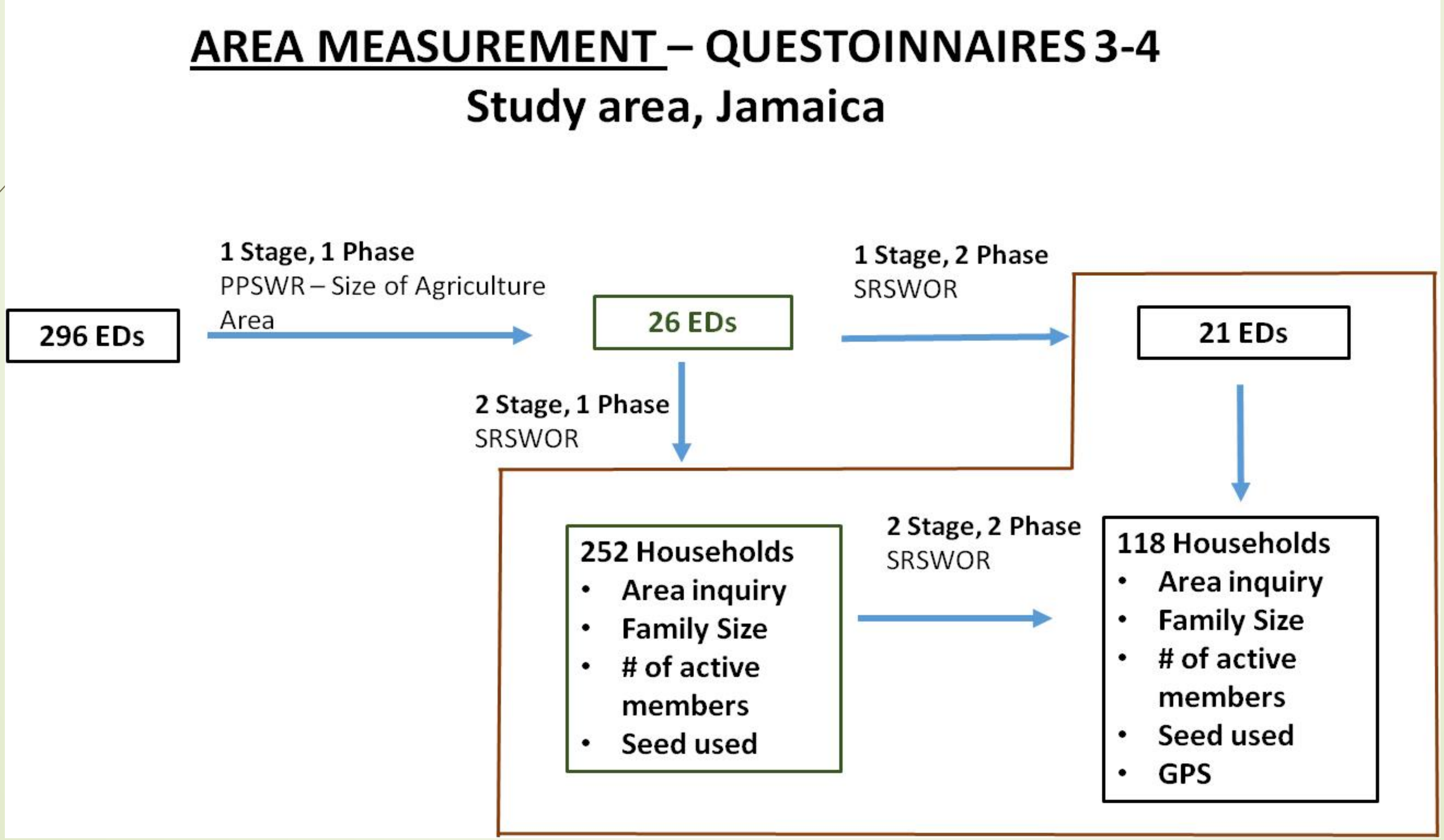
- Mixed and continuous cropping situation in the country
    - **Ministry of Agriculture and Fisheries (MoAF), Government of Jamaica identified a study area comprising of five Extension Areas (EAs).**
    - **Main food crops grown in the study area are Irish potato, sweet potato and yellow yam.**

Results & Discussions - Jamaica

Sampling Frame

- **The five extension areas in the study area were considered as strata**
- **There were 38, 78, 30, 109 and 41 EDs in first, second, third, fourth and fifth stratum**
- **From 296 EDs , a sample of 26 EDs was selected by PPSWR with agricultural area as size variable and completely enumerated**

Sampling design used for Crop Yield Estimation when Area Frame is Available



Description of Samples for Data Collection

Extension Area	Number of Enumeration Districts	Number of surveyed EDs for listing	Number of eligible HHs	Number of Selected Enumeration Districts/ HHs/parcels							
				Enumeration Districts				HHs/parcels			
				Area		Yield		Area (HHs)		Yield (parcels)	
				Enquiry	GPS	Assessment	CCE etc.	Enquiry	GPS	Assessment	CCE
Albert Town	38	8	165	4	4	3	2	46	28	28	17
Christiana	78	10	315	8	5	4	4	85	29	22	19
Comfort Hal	30	6	87	4	4	3	2	25	16	8	4
Lowe River	109	9	137	6	4	4	4	70	30	29	24
Warsop	41	10	437	4	4	4	2	26	15	14	6
Total	296	43	1141	26	21	18	14	252	118	101	29

Table 2.3.5.1: Extension Area wise, Number of Enumeration Districts, surveyed Enumeration Districts, selected Enumeration Districts and selected Households within selected Enumeration Districts for data collection on crop area and yield in different phases.

Period & Method of Data Collection

Period of Data Collection

- Period of data collection was March to July, 2016.
- The listing exercise was completed in the month of March, 2016. Data collection was carried out during April to July, 2016.

Method of Data Collection

- Both PAPI and CAPI methods used for data collection
- Listing exercise carried out using CAPI

Results & Discussions

GPS measurements ranged from 0.06 ha to 6.50 ha, while by farmer's inquiry the variation was from 0.1 ha to 8.0 ha. The parcel area in the study area in Jamaica is bigger than parcel area in Gunung Kidul district of Indonesia and Musanze district of Rwanda.

Percent relative difference between area by farmers' inquiry & GPS decreases as the area by farmers' inquiry increases

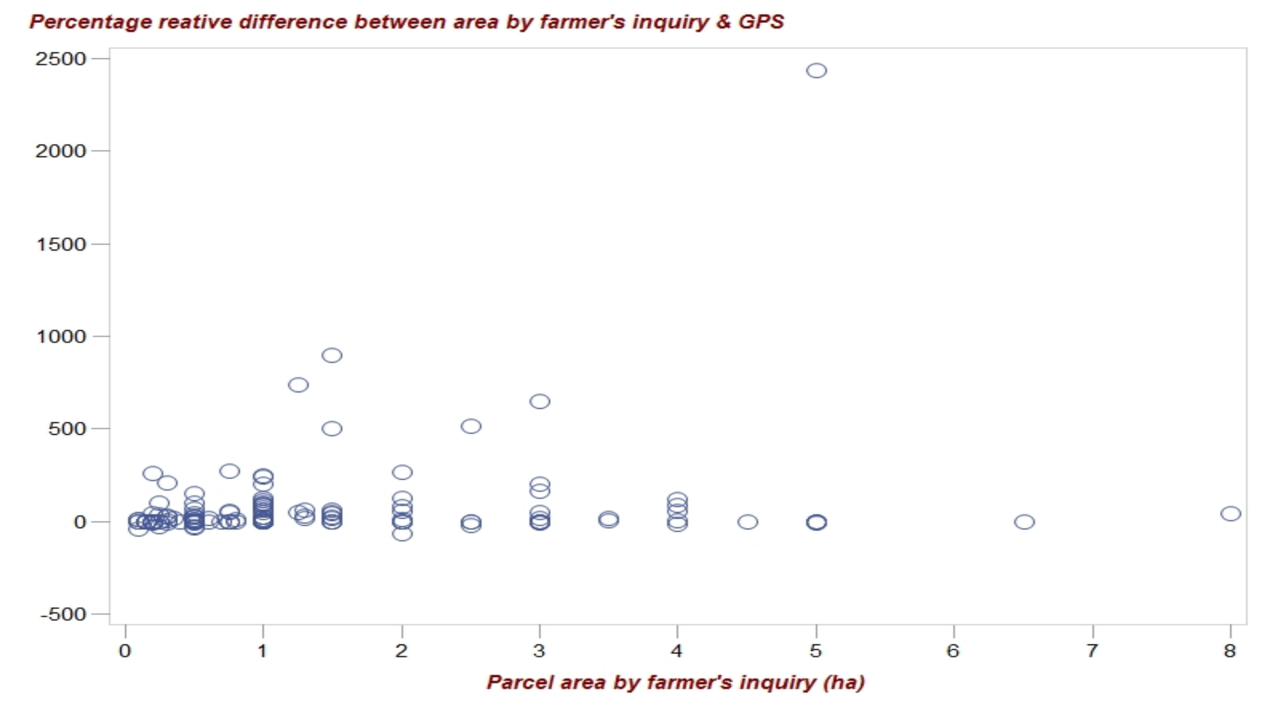


Figure 2.5.3.8.1: Percentage relative difference between the parcel area recorded by farmer inquiry and GPS in the study area in Jamaica

Results & Discussions

The sown area measured through GPS varied from 0.03 ha to 5.56 ha while by farmer's inquiry the variation was from 0.2 ha to 6.0 ha.

Percent relative difference between sown area by inquiry & GPS decreases as the sown area by farmers' inquiry increases

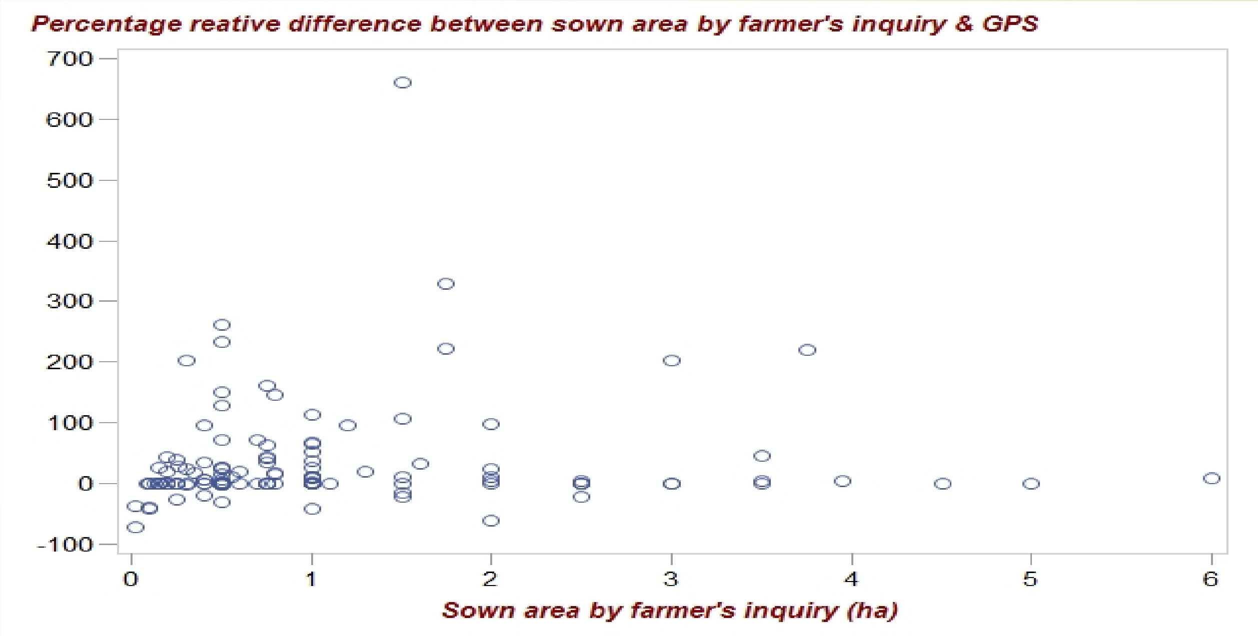


Figure 2.5.3.8.2: Percentage relative difference between the sown area recorded by farmer inquiry and GPS in the study area in Jamaica

Results & Discussions

Extension area	Total number of ED (N_h)	Large sample (Inquiry)	Small sample (GPS)
Albert Town	38	4	4
Christiana	78	8	5
Comfort Hall	30	4	4
Lowe River	109	6	4
Warsop	41	4	4
	296	26	21

Table 2.3.8.1: Extension area wise total number of enumeration districts along with enumeration district selected for inquiry based and GPS based crop area measurement

Results & Discussions

Extension area	Number of surveyed HHs for listing	Number of eligible HH	Number of Selected HHs/parcels			
			Area (HH)		Yield (Parcel)	
			Large sample (Inquiry)	Small sample (GPS)	Large sample (5A and eye estimate)	Small sample (CCE, SHU*)
Albert Town	165	165	46	28	5,5	5,5
Christiana	438	437	85	29	14,13	13,14
Comfort Hall	137	137	25	16	2,2	2,2
Lowe River	316	315	70	30	5,4	5,5
Warsop	87	87	26	15	4,4	4,3
TOTAL	1143	1141	252	118	30,28	29, 29

Table 2.3.8.2: Extension area wise number of households observed for listing, eligible households, and number of households selected for area enumeration and parcels selected for crop yield estimation.

Results & Discussions

Average relative difference in parcel area and sown area between GPS and farmer inquiry decreases as parcel area and sown area increase. The difference between GPS and farmer inquiry are narrower for sown area.

Area size class	##parcel	Average difference in Parcel area and farmers inquiry	Average percentage relative difference in Parcel area and farmers inquiry	Average difference in Sown area and farmers inquiry	Average percentage relative difference in Sown area and farmers inquiry
Over all	138	0.29	75.77	0.12	31.53
1	98	0.31	99.54	0.14	38.66
2	16	0.50	36.29	0.10	15.38
3	11	0.38	16.02	0.05	9.43
4	13	-0.18	-4.30	0.07	17.87

Table 2.3.8.3: Category-wise Average relative difference between parcel area measurement through GPS and farmer inquiry as well as Average relative difference between sown area measurement through GPS and farmer inquiry

Correlation matrix between different variables

The highest correlation is found between parcel area by GPS and enquiry based parcel area followed by total number of family members and active family members.

Variable	Total family member	Active family member	Area by inquiry	Area by GPS
Total family member	1.00	0.94	0.80	0.74
Active family member	0.94	1.00	0.90	0.79
Area by inquiry	0.80	0.90	1.00	0.99
Area by GPS	0.74	0.79	0.99	1.00

Table 2.3.8.4: Correlation matrix indicating correlation between total family members, active family members, parcel area by inquiry and parcel area by GPS

Contd…

As in case of parcel area, in case of sown area also high correlation exists between different variables. The trends in sown area are same as in case of parcel area.

Variable	Total family member	Active family member	Sown area by inquiry	Sown area by GPS
Total family member	1.00	0.94	0.78	0.70
Active family member	0.94	1.00	0.84	0.74
Sown area by inquiry	0.78	0.84	1.00	0.99
Sown area by GPS	0.70	0.74	0.99	1.00

Table 2.3.8.5: Correlation matrix indicating correlation between total family members, active family members, crop area sown by inquiry and crop area sown by GPS.

Estimates of Crop Area along with their Percentage Standard Errors

The percentage standard errors are least in the case of double sampling regression estimator involving GPS based crop area as study variable and farmer inquiry based crop area as auxiliary variable for all the three crops.

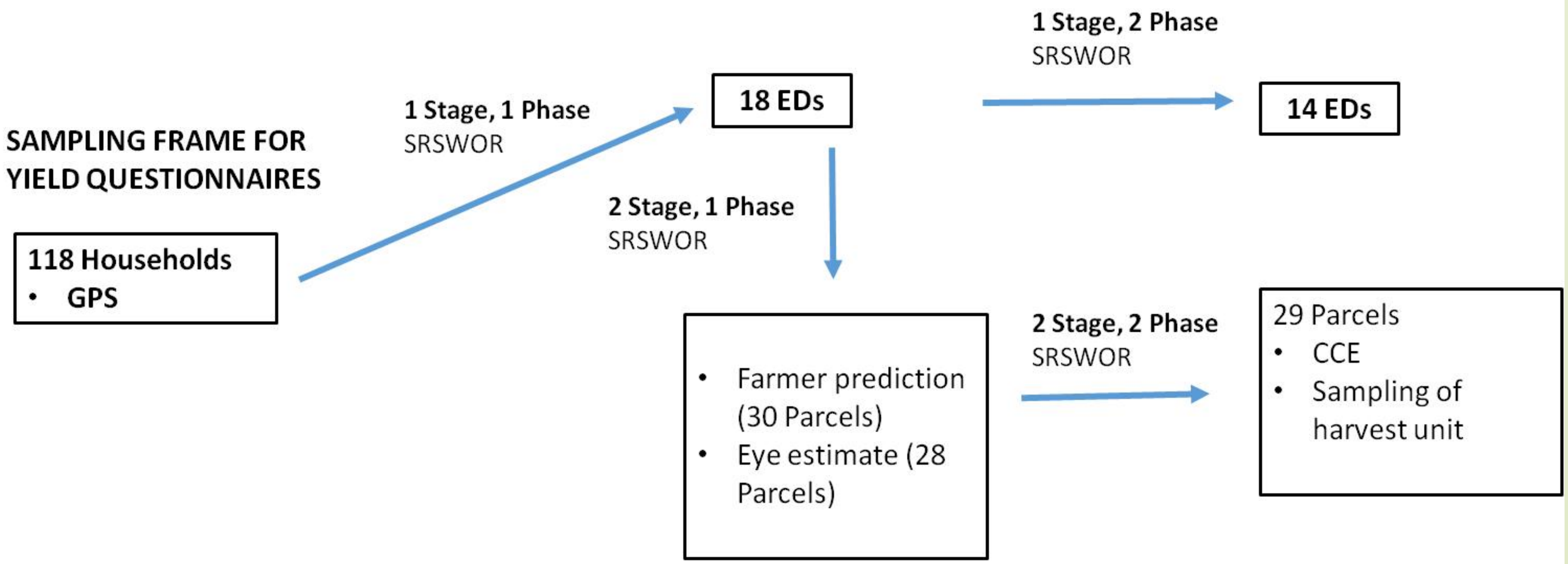
Crop Name	Crop wise sample sizes of EDs	Crop Area Estimates							
		By GPS only		Using total number of family members as auxiliary variables along with % SE		Using active family members as auxiliary variables along with % SE		Using Crop Area by inquiry as auxiliary variables along with % SE inquiry	
		$\hat{Y}_{FM}$ (ha)	%SE	$\hat{Y}_{GPS}$ (ha)	$\hat{Y}_{AFM}$ % SE	$\hat{Y}_{INQ}$ (ha)	% SE	(ha)	% SE
Irish Potato	5	88.88	60.78	69.99	76.28	77.21	69.34	300.06	14.63
Sweet Potato	5	125.69	73.49	138.53	66.33	143.94	63.86	259.88	35.39
Yellow Yam	18	3160.16	47.91	2601.63	57.96	2898.1	51.99	3460.21	43.39

Table 2.3.8.6: GPS based linear estimates as well as double sampling regression estimates using total number of family members, active family members and inquiry based crop area as auxiliary variables along with the percentage standard errors.

Sampling design for crop yield estimation for mixed cropping using household approach

PRODUCTION MEASUREMENT – QUESTOINNAIRES 5A,5B, CCE1,CCE2

Study area, Jamaica



Crop Yield Estimation

Estimator based on farmer assessment of crop produce beats the other estimators in terms of the criterion of percentage standard error of the estimator.

Crop name	CCE		Farmer assessment (5A)		Eye estimate		Sampling of harvest unit	
	Est. (kg/ha)	% SE	Est. (kg/ha)	% SE	Est. (kg/ha)	% SE	Est. (kg/ha)	% SE
Yellow Yam	27502.04	27.27	13716.28	27.73	16247.37	31.22	11010.00	47.50
Sweet Potato	7603.11	12.86	2045.04	4.10	2448.00	5.24	2224.14	0.67

Table 2.3.8.7 Estimates of crop yield by CCE, Farmer assessment (5A),Eye Estimate on day of harvesting and sampling of harvest unit along with percentage standard error using simple linear estimator.

Optimum Values of Sample Sizes

The optimum values of sample sizes are smallest when measurements are taken by farmer assessment
The method of sampling of harvest unit exhibits maximum variability.

Crop Name	CCE	Farmer assessment (5A)	Eye Estimate	Sampling of harvest unit
Yellow Yam	121	118	125	366
Sweet Potato	13	2	2	2

Table 2.3.8.8: Optimum values of sample size for crop yield estimation at 5% CV

Challenges Encountered

- **Problems in data collection by CAPI**
  Trained supervisors/manpower at all levels are required for data collection through CAPI.
- **The process of substitution of sampling units in data collection through CAPI was found to be cumbersome.**
- **Besides, the survey in question involved determination of crop area and crop yield through alternative methods.**
- **Multiple visits were required for collection of data.**
- **Created implementation problems in CAPI resulting in data as well as time loss in data collection.**

Challenges Encountered

- **Insufficient survey period for data collection**
  For adequately capturing continuous cropping pattern one full agricultural year is required. However, in this study data collection was restricted to five months including listing of households. As a result, issue of continuous cropping could not be addressed properly in Jamaica
- **Problems of area measurement by GPS**
  It was proposed to collect crop area data by several methods including GPS. Many parts in study area lay in hilly areas. The satellite connection in these areas was limited. This affected functioning of the GPS instruments.

Challenges Encountered

- **Supervision and monitoring**
  Two visits were planned for the ICAR-IASRI team in the field testing countries for imparting training and another visit for supervision of field work. An additional visit was felt necessary for proper supervision of field work.

Challenges of sampling methodology

Although utility of two-stage sampling designs are well known that they are useful when frame of ultimate units is not available, the precision of an estimator suffers under two-stage sampling designs. Thus, larger sample sizes are needed to keep the precision intact

- The selection of an overall sample, as in case of proposed methodology, is very
- convenient in that it does away with the requirement of mixture by mixture sampling ,
- but it means that some of the mixtures/pure stands may be over-represented while some
- others may be under-represented.

- The use of overall multipliers in domain estimation is practically convenient for large scale surveys in the sense that it does away with the requirement of using mixture by mixture multipliers, which more often than not are unavailable. But the use of overall multipliers reduces the precision of the estimators.

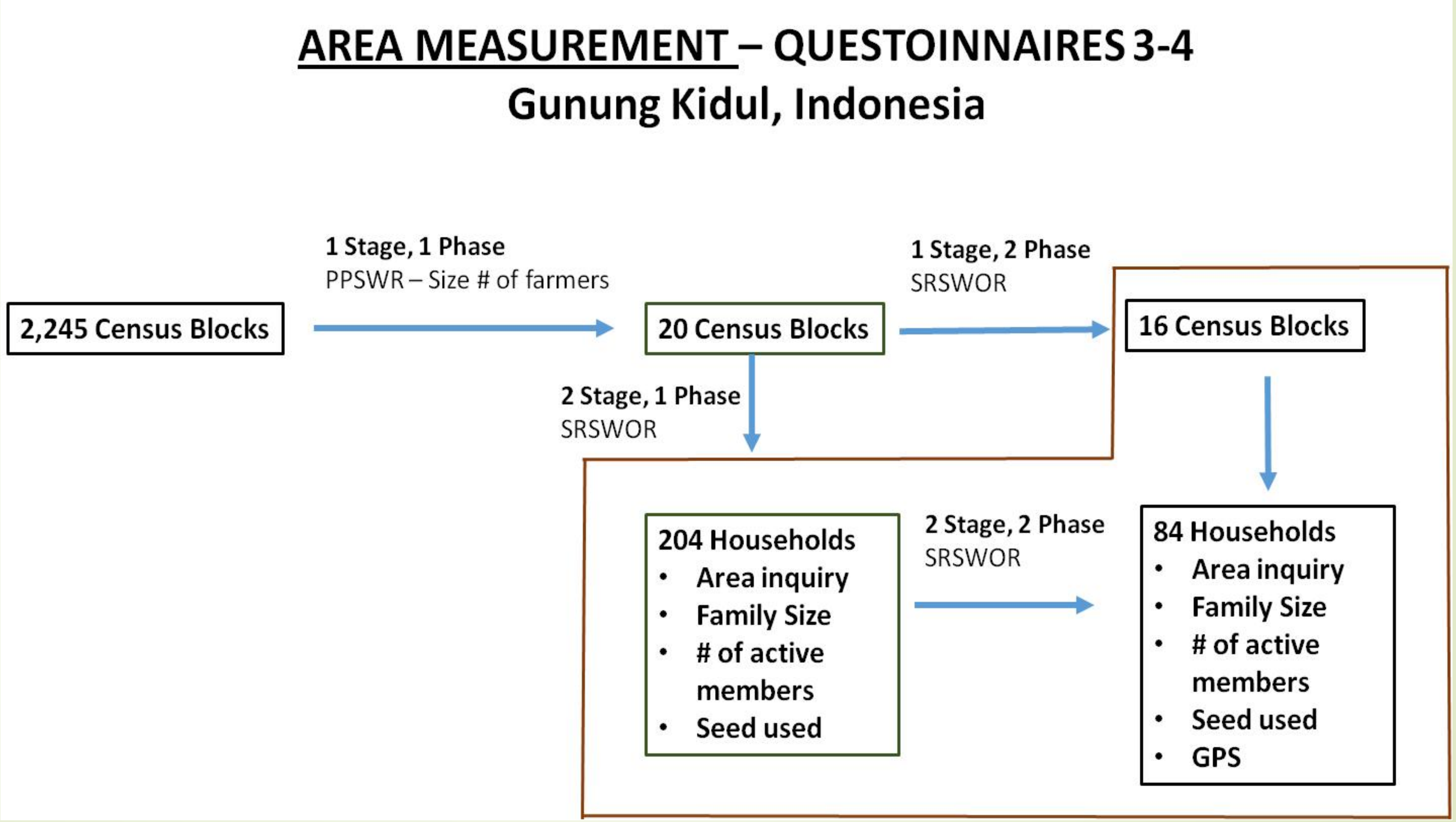
Contd…

- Local statisticians may find the stratified two-stage sampling design with two phases bit
- involved. Similarly, they may have difficulty in selection of the sample as per the design.
- Intensive training is required on the sampling design as well as selection of sample through
- sofware

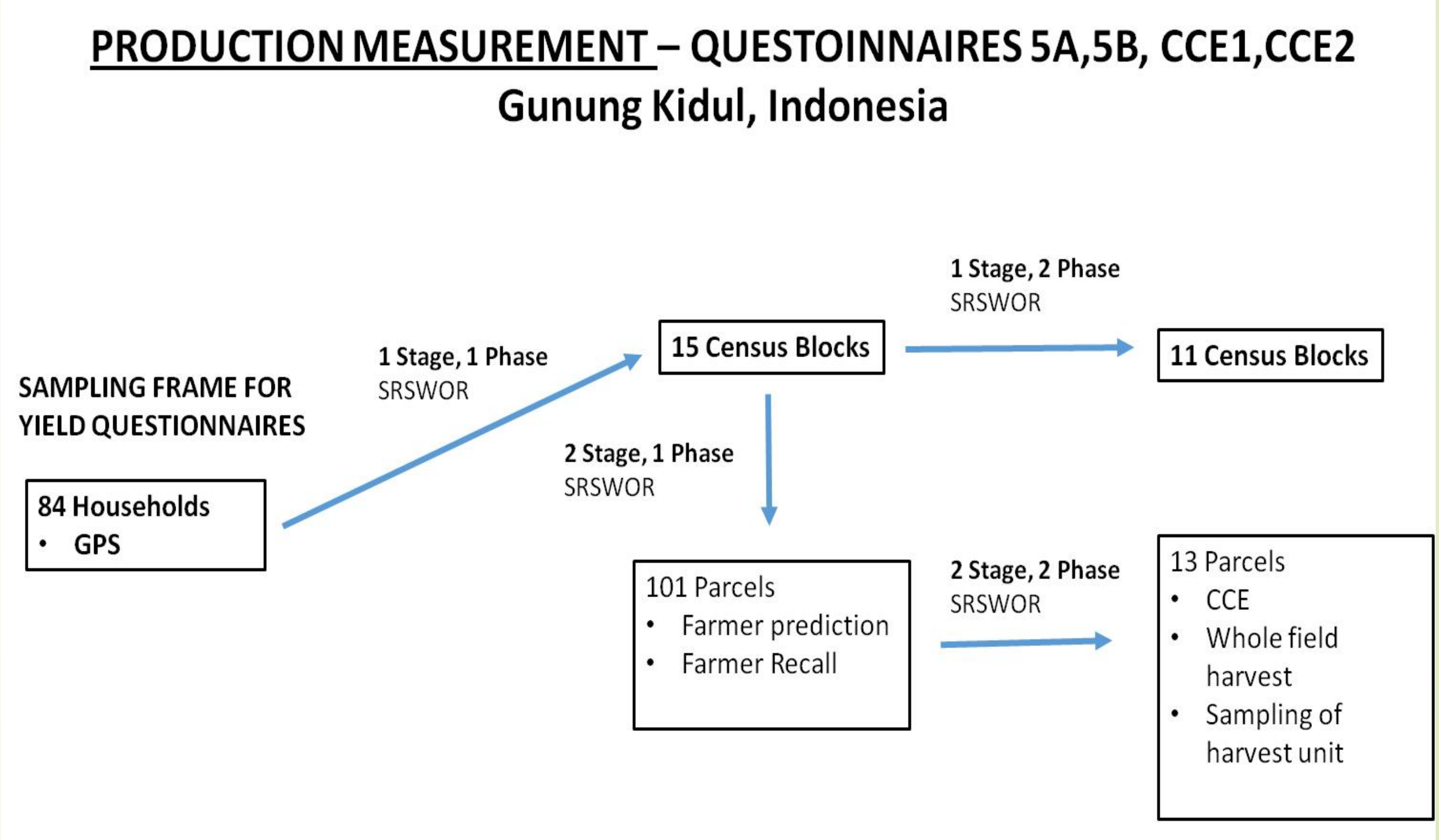
Although, as in case of proposed methodology also, five visits are absolute minimum in any survey for listing as well as data collection on crop areaand yield estimation, the clubbing of data on different variables in a single visit in the case of proposed methodology means the local statisticians will be required to get accustomed to such a mechanism of data collection.

The requirement of data collection through farmer prediction and farmer recall within 15 days period implies considerable planning in the organization of the survey so that enumerators are able to collect data as per requirement of methodology

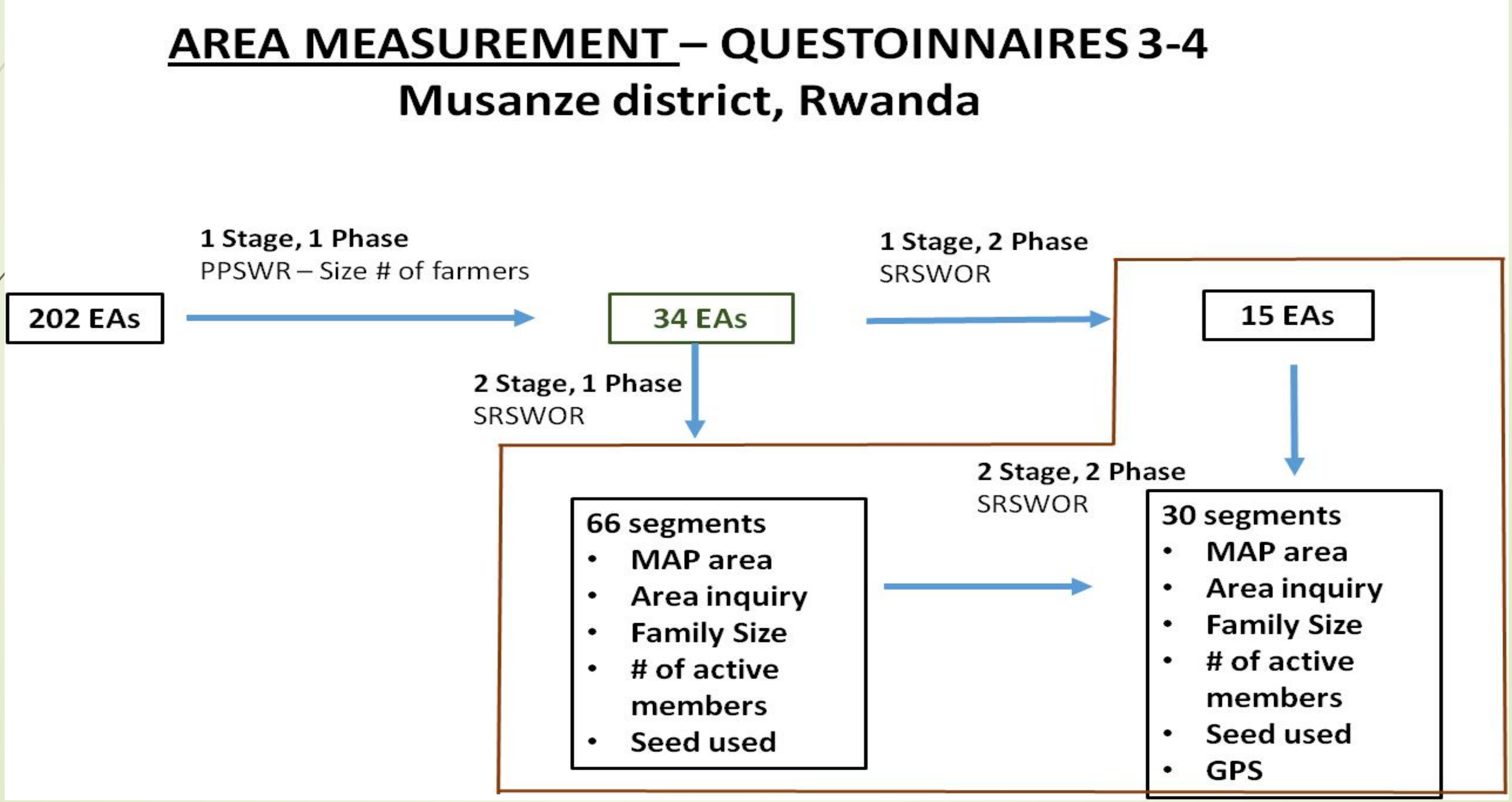
Sampling design for crop area estimation for mixed cropping using household approach



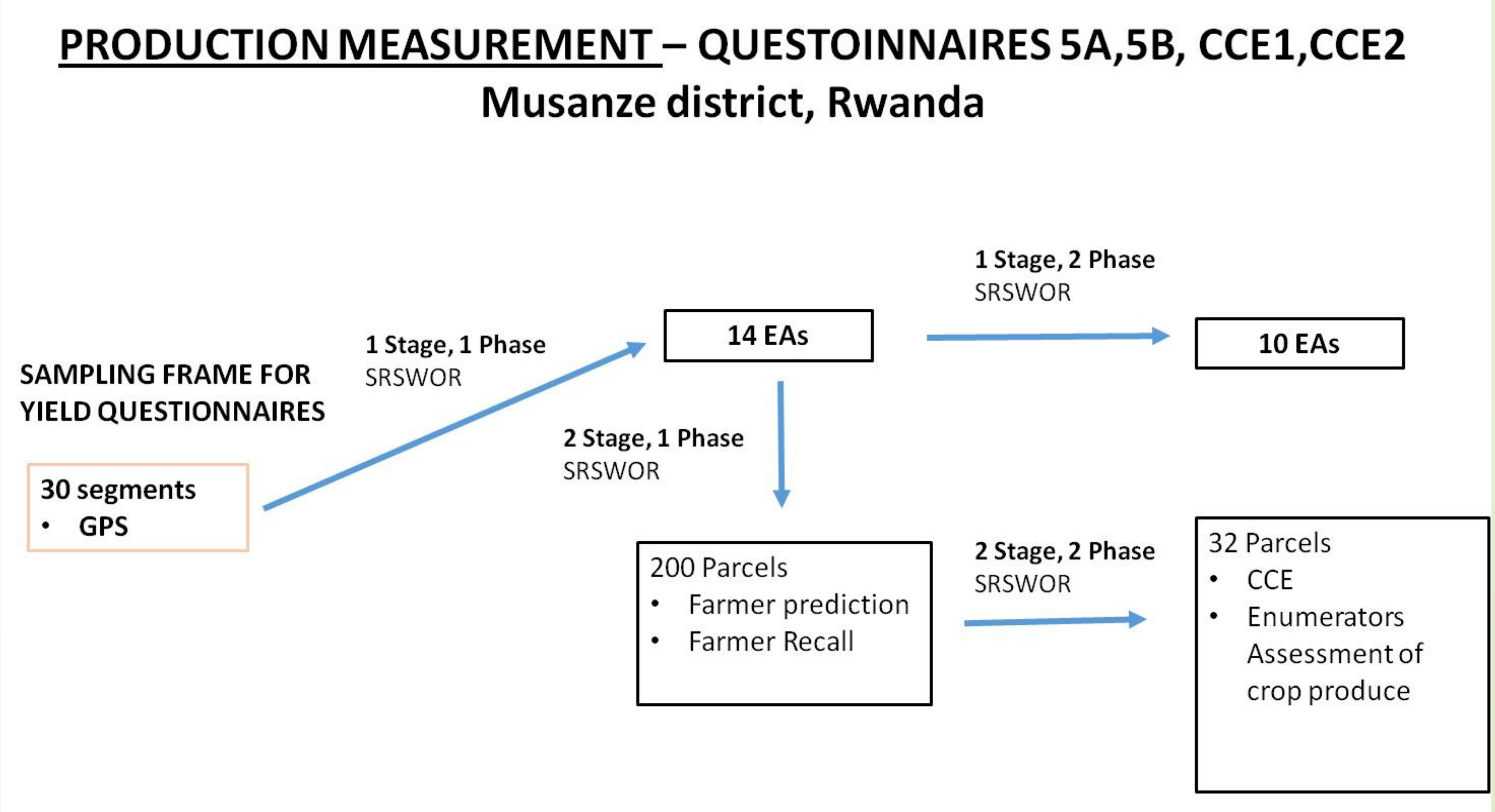
‘Sampling design for crop yield estimation for mixed cropping using household approach



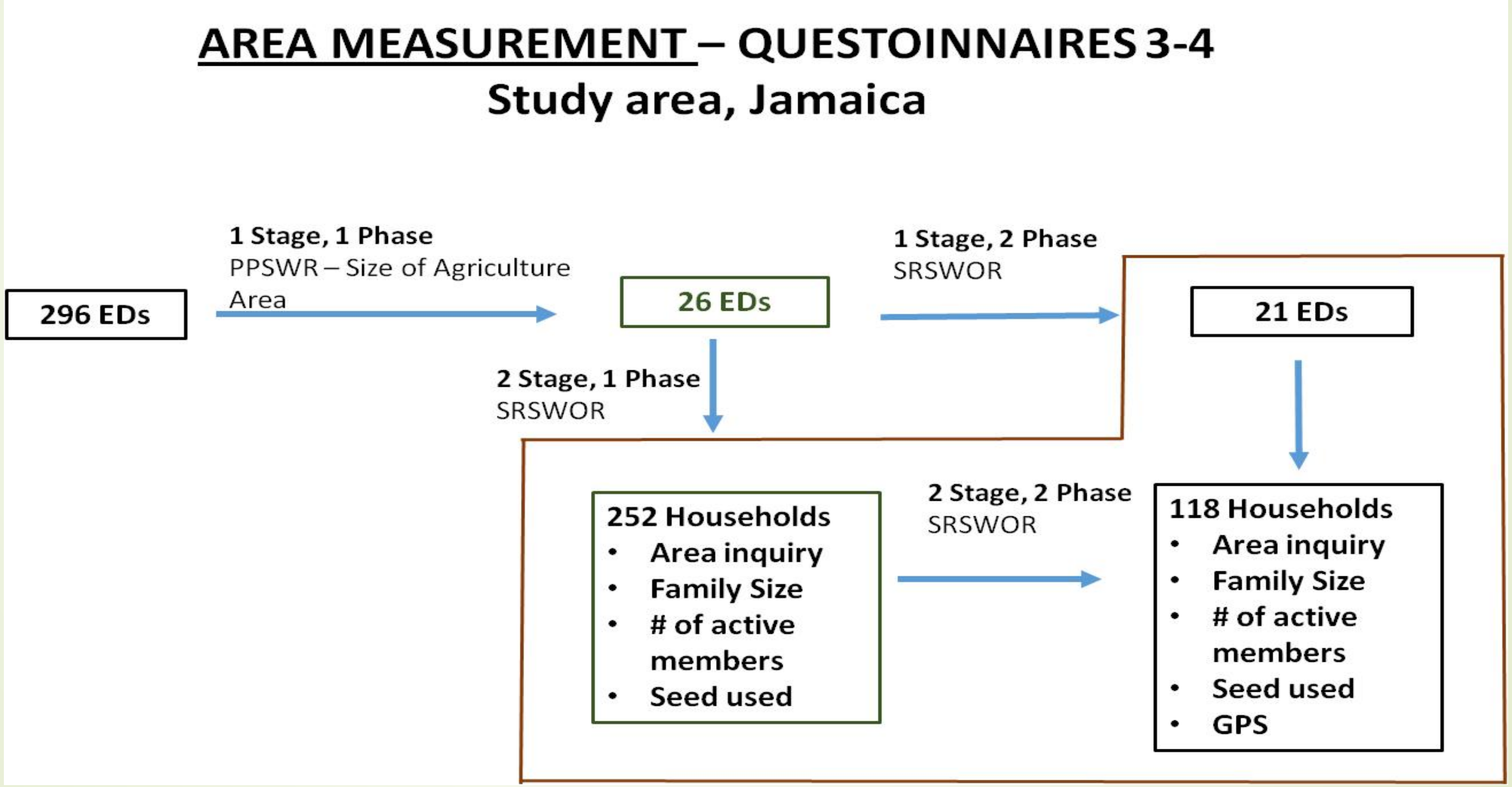
Sampling design for crop area estimation for mixed cropping when Area Frame is available



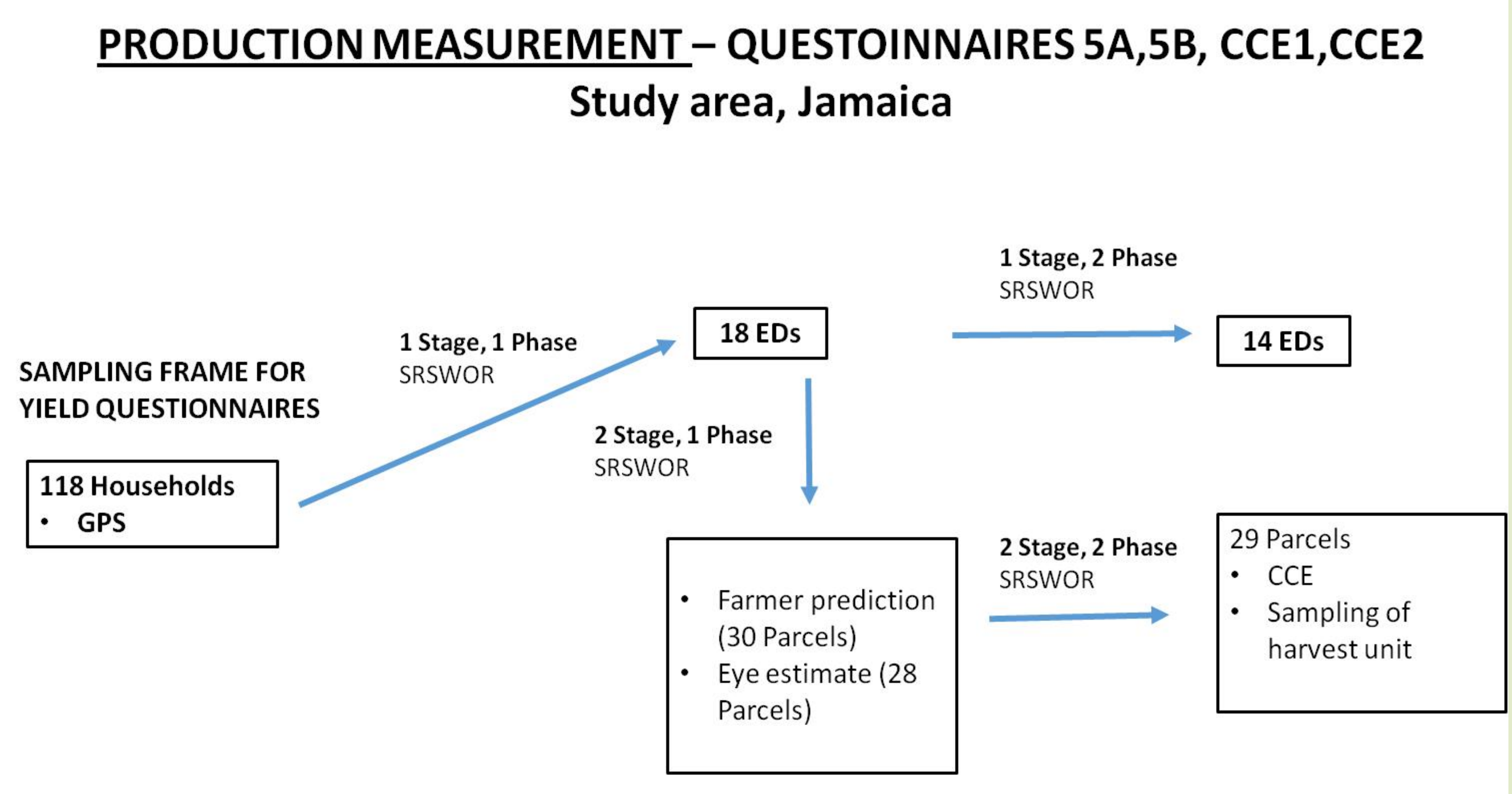
Sampling design for crop yield estimation for mixed cropping when area frame is available



Sampling design for crop area estimation for mixed cropping using household approach



Sampling design for crop yield estimation for mixed cropping using household approach



Estimation Procedure (Household Approach)

Crop Area Estimation

H = sub-districts which can be considered as *H* strata in a district.

N_h= number of ED/Census Block (CB) in the h^{th} sub-district (stratum)which are considered as the PSUs, h=1,…,H

N_h is known.

M_{hi} = number of households (SSU) in each ED/CB.

T_{hij} = No. of parcels (USU) in the j-th Household of i-th ED/CB where different forms of crop mixture e.g. as a pure stand, mixture-1, mixture-2 etc. are grown

The different crop mixtures are taken as domains of the study. We assume that in each h^{th} sub-districts, *D* different crop mixtures are being grown as pure stand, mixture-1, mixture-2 etc.

Estimation Procedure (Household Approach)

There would be $\{U_{h1}, \ldots, U_{hd}, \ldots, U_{hD}\}$ domains in the h^{th} stratum, U_h

Although, number of EDs/CBs (N_h), number of households (M_{hi}) and number of parcels (T_{hij}) in each sub-district are known (for the selected EDs/CBs), for a specific d^{th} form of mixture, number of EDs/CBs (N_{hd}), number of households (M_{hid}) and number of parcels (T_{hijd}) are generally unknown

Let y_{hijk} = crop area of k^{th} parcel (USU) in the j^{th} household (SSU) of the i^{th} ED/CB (PSU) in the h^{th} sub-district (stratum) of a district

Estimation Procedure (Household Approach)

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

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

The population total based on all the domains is given as

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

Sampling design for estimation of crop area at district level using household approach is stratified two stage cluster sampling design with two phases in each stages of sampling. Suppose, in the first phase, a PPSWR sample of size n'_h is drawn from N_h villages/ED/CB (PSU). The probability of selecting i^{th} village/ED/CB in the h^{th} stratum is taken n_h $Z_{hi} = X_{hi}/X_h$

where X may be taken as total agricultural land/no. of farmers.

m_{hi}' = no. of households selected by SRSWOR design in the second stage of first phase and all T_{hij} parcels in selected household are completely enumerated for collecting auxiliary information regarding the parcel like seed used, farmers' assessment etc. etc.

Estimation Procedure (Household Approach)

- n_{hd}' = no. of villages out of selected n_h' villages are following the specific d^{th} mixture.
- m_{hid}' = no. of households out of m_{hi}' selected households growing d-th mixture in their T_{hijd} total parcels.
- In the second phase,

  n_h = no. of EDs/CB selected from n_h' initially selected EDs/CB by SRSWOR.

  m_{hi}-= number of households selected in each selected ED/CB by SRSWOR design and all the T_{hij} parcels in selected household are completely enumerated.
- The areas of each of the parcels growing crop mixtures of these sampled households are measured by GPS.

Estimation Procedure (Household Approach)

The area for component crops are obtained by apportioning using seed rate information.

n_{hd} = number of villages out of n_h villages growing the d-th mixture

m_{hid} = number of households out of m_{hi} households growing the d-th mixture

T_{hijd} = number of parcels out of T_{hi} parcels growing the d-th mixture

Aim is to estimate the total crop area under a specific crop(Y) as well as under different mixtures (Y_d), $d=1,2,\ldots,D$

Let x_{hijk} = auxiliary information (e.g. seed used, farmers' assessment etc.) corresponding to k^{th} parcel (SSU) of the j^{th} selected household in the i^{th} ED/CB (PSU) of the h^{th} sub-district (stratum)

Whereas, y_{hijk} = Crop area measured by GPS method of k^{th} parcel (USU) in the j^{th} household (SSU) of the i^{th} villages/CB (PSU) in the h^{th} sub-district (stratum) of a district.

A regression estimator of the total area for d^{th} mixture under the stratified two phase two stage cluster sampling design is

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

Estimation Procedure (Household Approach)

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

$$\hat{X}_d = \sum_{h=1}^{H} \frac{1}{n_h} \sum_{i=1}^{n_{hd}} \frac{\hat{X}_{hid}}{z'_{hi}}, \quad \hat{X}'_d = \sum_{h=1}^{H} \frac{1}{n'_h} \sum_{i=1}^{n_{hd}} \frac{\hat{X}'_{hid}}{z'_{hi}},$$

$$\hat{Y}_{hid} = \frac{M_{hi}}{m_{hi}} \sum_{j=1}^{m_{hid}} \sum_{k=1}^{T_{hijd}} y_{hijk}, \quad \hat{X}_{hid} = \frac{M_{hi}}{m_{hi}} \sum_{j=1}^{m_{hid}} \sum_{k=1}^{T_{hijd}} x_{hijk}, \quad \hat{X}'_{hid} = \frac{M_{hi}}{m'_{hi}} \sum_{j=1}^{m'_{hid}} \sum_{k=1}^{T_{hijd}} x_{hijk}.$$

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

$$b_{A3d} = c_{A3xyd} / c_{A3xxd},$$

Estimation Procedure (Household Approach)

$$c_{A3xyd} = \sum_{h=1}^{H}\left(\frac{1}{n_h} - \frac{1}{n_h{'}}\right)\left(p_{hd}\, s_{bxyhd} + p_{hd}\, q_{hd}\, \hat{X}_{hd}\hat{Y}_{hd}\right),$$

$$c_{A3xxd} = \sum_{h=1}^{H}\left(\frac{1}{n_h} - \frac{1}{n_h{'}}\right)\left(p_{hd}\, s_{bxhd}^{2} + p_{hd}\, q_{hd}\, \hat{X}_{hd}^{2}\right),$$

$$p_{hd} = \frac{n_{hd}}{n_h}, \quad q_{hd} = 1 - p_{hd},$$

$$s_{bxhd}^{2} = \frac{1}{n_{hd} - 1}\sum_{i=1}^{n_{hd}}\left(\frac{M_{hi}\, p_{hid}\, \overline{x}_{hid}}{z_{hi}^{'}} - \hat{X}_{hd}\right)^{2},$$

$$s_{bxyhd} = \frac{1}{n_{hd} - 1}\sum_{i=1}^{n_{hd}}\left(\frac{M_{hi}\, p_{hid}\, \overline{x}_{hid}}{z_{hi}^{'}} - \hat{X}_{hd}\right)\left(\frac{M_{hi}\, p_{hid}\, \overline{y}_{hid}}{z_{hi}^{'}} - \hat{Y}_{hd}\right),$$

$$\overline{x}_{hid} = \frac{1}{m_{hid}}\sum_{j=1}^{m_{hid}} x_{hij.d}, \quad \overline{y}_{hid} = \frac{1}{m_{hid}}\sum_{j}^{m_{hid}} y_{hij.d},$$

$$\hat{X}_{hd} = \frac{1}{n_{hd}}\sum_{i=}^{n_{hd}}\frac{M_{hi}\, p_{hid}\, \overline{x}_{hid}}{z_{hi}^{'}}, \quad \overline{y}_{hd} = \frac{1}{n_{hd}}\sum_{i=1}^{n_{hd}}\frac{M_{hi}\, p_{hid}\, \overline{y}_{hid}}{z_{hi}^{'}}.$$

Estimation Procedure (Household Approach)

An approximate estimate of variance of regression estimator $\hat{Y}_{lr3d}$ is given as

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

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

the estimator of percentage standard error of the proposed regression estimator of total area under d^{th} mixture $\hat{Y}_{lr3d}$ is given by

$$\%\, S\widehat{E}(\hat{Y}_{lr3d})=\frac{\sqrt{\hat{V}\left(\hat{Y}_{lr3d}\right)}}{\hat{Y}_{lr3d}}\times 100.$$

Estimation Procedure (Household Approach)

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

$$\hat{Y}_{c3} = \sum_{d=1}^{D*} \hat{Y}_{lr3d}$$

the estimator of percentage standard error of the proposed estimator of population total $\hat{Y}_{c3}$

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

$$\% \, \widehat{SE}(\hat{Y}_{c3}) = \frac{\sqrt{\hat{V}\left(\hat{Y}_{c3}\right)}}{\hat{Y}_{c3}} \times 100.$$

Crop Yield Estimation Household Approach

At the time of surveying the parcels for area enumeration, a list of parcels growing different mixtures is prepared in a ED/CB.

Let, m_{hid} = no. of parcels growing d^{th} mixture of the crop out of m_{hi} sampled parcels in the n_{hd} EDs/CBs of the h^{th} sub-districts,

In the first phase of sampling,

$n_h^{"}$ = no. of EDs/CBs selected by SRSWOR design from n_h EDs/CBs (PSU)

$m_{hid}^{"}$ = no. of parcels selected by SRSWOR for the d-th mixture from m_{hid} parcels within each selected ED/CB $\forall\ d = 1, 2, ..., D$

From the selected parcels, farmers' eye estimates of harvested yield of the crop under mixture is recorded.

x_{hijd} = the eye-estimated harvested yield of the crop in the j^{th} parcel growing the d^{th} mixture in the i^{th} ED/CB of the h^{th} sub district $d = 1, 2, ..., D$

Contd…

Let, x_{hijd} = eye-estimated harvested yield of the crop in the j^{th} parcel growing d^{th} mixture in the i^{th} ED/CB of the h^{th} sub district

In the second phase of sampling,

$n_h''' = \text{no. of}$ EDs/CBs selected by SRSWOR design from the n_h'' selected EDs/CBs and within each selected ED/CB for each mixture, d=1,2,…,D,

let, m_{hid}'''= no. of parcels selected from the m_{hid}'' first phase second stage units in the i^{th} ED/CB of the h^{th} sub district

Contd…

- In the final sample, it can be observed that out of n_h''' sample EDs/CBs, n_{hid}'' ∀d=1,2,…,D, EDs/CBs are following the d^{th} mixture,

Now, from all these m_{hid}''' parcels in the final sample, estimates of the harvested yields of a crop under a specific mixture are obtained by conducting CCE etc. for mixed cropping.

Contd...

- Let, x_{hijd} = harvested yield of the crop under d^{th} mixture from the j^{th} parcel in the i^{th} CB/ED of the h^{th} sub district

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

$$\overline{y}_{lr3d} = \overline{y}_{3d} + b_{3d}(\overline{x}'_{3d} - \overline{x}_{3d}) \quad (8)$$

Contd…

where,

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

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

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

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

Contd…

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

$$c_{3xxd} = \frac{1}{H^2}\sum_{h=1}^{H}\left[\left(\frac{1}{n_h'''} - \frac{1}{n_h''}\right)(p_{hd}s_{bxhd}^2 + p_{hd}q_{hd}\overline{x}_{hd}^2) + \frac{1}{n_h'''n_h''}\sum_{i}^{n_{hd}'''}\left(\frac{1}{m_{hid}'''} - \frac{1}{m_{hid}''}\right)s_{xhid}^2\right],$$

$$s_{bxhd}^2 = \frac{1}{n_{hd}''' - 1}\sum_{i=1}^{n_{hd}'''}(\overline{x}_{hid} - \overline{x}_{hd})^2,$$

$$s_{bxyhd} = \frac{1}{n_{hd}''' - 1}\sum_{i=1}^{n_{hd}'''}(\overline{x}_{hid} - \overline{x}_{hd})(\overline{y}_{hid} - \overline{y}_{hd}),$$

Contd…

$$s^2_{xhid} = \frac{1}{m'''_{hid} - 1} \sum_{j=1}^{m'''_{hid}} \left(x_{hijd} - \overline{x}_{hid} \right)^2,$$

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

$$p_{hd} = \frac{n'''_{hd}}{n'''_{h}}, \quad q_{hd} = 1 - p_{hd},$$

$$p_{hid} = \frac{m'''_{hid}}{m'''_{hi}}, \quad q_{hid} = 1 - p_{hid},$$

$$\overline{x}_{hid} = \frac{1}{m'''_{hid}} \sum_{j}^{m'''_{hid}} x_{hijd}, \quad \overline{y}_{hid} = \frac{1}{m'''_{hid}} \sum_{j}^{m'''_{hid}} y_{hijd},$$

Contd…

$$\overline{x}_{hd} = \frac{1}{n'''_{hd}} \sum_{i=}^{n'''_{hd}} \overline{x}_{hid}\,, \qquad \overline{y}_{hd} = \frac{1}{n'''_{hd}} \sum_{i=1}^{n'''_{hd}} \overline{y}_{hid}\,.$$

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

$$\hat{V}\left(\overline{y}_{lr3d}\right) = \frac{1}{H^2}\sum_{h=1}^{H}\Bigg[\left\{\left(\frac{1}{n'''_h} - \frac{1}{N_h}\right)\left(p_{hd}s^2_{byhd} + p_{hd}q_{hd}\overline{y}^2_{hd}\right) + \frac{1}{n'''_h N_h}\sum_{i=1}^{n'''_{hd}}\left(\frac{1}{m'''_{hid}} - \frac{1}{M_{hid}}\right)s^2_{yhid}\right\}$$
$$- r^2_{3d}\left\{\left(\frac{1}{n'''_h} - \frac{1}{n''_h}\right)\left(p_{hd}s^2_{byhd} + p_{hd}q_{hd}\overline{y}^2_{hd}\right) + \frac{1}{n'''_h n''_h}\sum_{i=1}^{n'''_{hd}}\left(\frac{1}{m'''_{hid}} - \frac{1}{m''_{hid}}\right)s^2_{yhid}\right\}\Bigg], \qquad (9)$$

Contd…

and,

$$\% \widehat{SE}(\bar{y}_{lr3d}) = \frac{\sqrt{\hat{V}(\bar{y}_{lr3d})}}{\bar{y}_{lr3d}} \times 100. \qquad (10)$$

where,

$$s^2_{byhd} = \frac{1}{n'''_{hd} - 1} \sum_{i=1}^{n'''_{hd}} \left(\bar{y}_{hid} - \bar{y}_{hd}\right)^2,$$

$$s^2_{yhid} = \frac{1}{m'''_{hid} - 1} \sum_{j=1}^{m'''_{hid}} \left(y_{hijd} - \bar{y}_{hid}\right)^2,$$

$$r^2_{3d} = \frac{c^2_{3xyd}}{c_{3xxd}\, c_{3yyd}},$$

Contd...

and, C_{3yyd} is in same functional form as C_{3xxd} defined earlier

The estimator of yield for a specific crop *c* based on all the domains at district level along with its approximate estimator of the variance is given by

$$\hat{\bar{Y}}_{c3} = \frac{1}{D^*} \sum_{d=1}^{D^*} \bar{y}_{lr3d} \qquad (11)$$

$$\hat{V}\left(\hat{\bar{Y}}_{c3}\right) = \frac{1}{D^{*2}} \sum_{d=1}^{D^*} \hat{V}\left(\bar{y}_{lr3d}\right) \qquad (12)$$

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

Contd…

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

$$\% \widehat{SE}\left(\hat{\bar{Y}}_{c3}\right) = \frac{\sqrt{\hat{V}\left(\hat{\bar{Y}}_{c3}\right)}}{\hat{\bar{Y}}_{c3}} \times 100. \qquad (13)$$

2.0 Area Fame Approach

- *Estimation procedure for crop area under mixed cropping when Area frame available*

Let, H = sub-districts within a district which can be considered as H strata

N_h = no. of EAs in the h^{th} sub-district (stratum) which are considered as the PSUs, h=1,…,H

Total number of EAs in each sub-districts (N_h) are known

Let, M_{hi} = no. of segments (SSU) in h-th sub-district of i-th EA parcels (USU)

$\boldsymbol{T_{hij}}$ = no. of parcels in the j-th segment of i-th EA of h-th sub-district

Contd…

In these parcels crops are grown in different forms of mixture e.g. as a pure stand, mixture-1, mixture-2 etc

The different crop mixtures are taken as domains of the study

Let in h^{th} sub-districts, D different crop mixtures are being grown as pure stand, mixture-1, mixture-2 etc

Thus, there would be $\{U_{h1}, \ldots, U_{hd}, \ldots, U_{hD}\}$ domains in the h^{th} stratum

Let,

y_{hijk} = crop area of k^{th} parcel (USU) in the j^{th} segment (SSU) of the i^{th} EA (PSU) in the h^{th} sub-district (stratum)

Contd...

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

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

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

the population total based on all the domains is given

Contd…

- The proposed sampling design for estimation of crop area at district level is stratified two stage cluster sampling design with two phases in each stages of sampling using Area Frame approach

- Let, in the first phase,

- n_h' = no. of EAs selected from N_h EAs (PSU) by probability proportional to the size with replacement (PPSWR) design in the h^{th} stratum.

Contd…

The probability of selecting i^{th} EA in the h^{th} stratum is taken z'_{hi} = X_{hi}/X_h, where X may be taken as total agricultural land.

In the second stage of first phase,

m_{hi}' = no. of segments selected by SRSWOR design from the set of M_{hi} segments within each selected EA

T_{hij} = total no. of parcels, in a selected tract, from the j-th segment of i-th EA in the h-th strata growing crop mixtures for collecting auxiliary information regarding the parcel like seed used, farmers' assessment etc. , j=1,…m_{hi}' .

Contd…

- Let, n_{hd}' = no.of EAs in which d-th mixture is grown out of n_h' selected EAs.
- T_{hijd} = no. of parcels in which d-th mixture in grown in j-th segment of i-th EA of h-th sub-district
- m_{hid}'=no. of segments in which d-th mixture is grown in the selected ith EA of h-th sub-district

- In the second phase of sampling, let
- n_h = no. of EAs selected from n_h' initially selected EAs (PSU) by SRSWOR.
- In each of these selected EAs, a sub-sample of
- m_{hi} = no. of segments selected by SRSWOR design and all the T_{hij} parcels in selected EA are completely enumerated
- The parcel area are measured by GPS

Contd…

- Data collection was made using the questionnaires given in Annexure B of Field Test Protocol document
- The area for component crops were obtained by apportioning using seed rate information
- Let n_{hd} = no. of EAs in which d-th mixture is grown out of n_h EAs in the h-th sub-district
- m_{hid} = no. of segments out of m_{hi} segments growing d-th mixture in i-th EA of h-th sub-district
- T_{hijd} = no. of parcels growing d-th mixture in j-th segment of i-th EA of h-th sub-district

Contd…

- In current scenario our aim is to estimate the total crop area under a specific crop(Y) as well as under different mixtures (Y_d), d =1,2,…, D.
- Let, x_{hijk} = auxiliary information (e.g. seed used, farmers' assessment etc.) corresponding to k^{th} parcel (SSU) of the j^{th} selected segment in the i^{th} EA (PSU) in the h^{th} sub-district (stratum), whereas,
- y_{hijk} = crop area of the parcel measured by GPS and Farmer inquiry.

Contd…

- A regression estimator of the total area under a d^{th} mixture under the stratified two phase two stage cluster sampling design can be formed as

Where, $$\hat{Y}_{lr2d} = \hat{Y}_d + b_{A2d}(\hat{X}_d{}' - \hat{X}_d) \quad (10)$$

$$\hat{Y}_d = \sum_{h=1}^{H} \frac{1}{n_h} \sum_{i=1}^{n_{hd}} \frac{1}{z'_{hi}} \frac{M_{hi}}{m_{hi}} \sum_{j=1}^{m_{hid}} \sum_{k=1}^{T_{hijd}} y_{hijk} = \sum_{h=1}^{H} \frac{1}{n_h} \sum_{i=1}^{n_{hd}} \frac{1}{z'_{hi}} \frac{M_{hi}}{m_{hi}} \sum_{j=1}^{m_{hid}} y_{hij.d} = \sum_{h=1}^{H} \frac{1}{n_h} \sum_{i=1}^{n_{hd}} \frac{\hat{Y}_{hid}}{z'_{hi}},$$

$$\hat{X}_d = \sum_{h=1}^{H} \frac{1}{n_h} \sum_{i=1}^{n_{hd}} \frac{\hat{X}_{hid}}{z'_{hi}}, \quad \hat{X}'_d = \sum_{h=1}^{H} \frac{1}{n'_h} \sum_{i=1}^{n_{hd}} \frac{\hat{X}'_{hid}}{z'_{hi}}$$

$$\hat{Y}_{hid} = \frac{M_{hi}}{m_{hi}} \sum_{j=1}^{m_{hid}} \sum_{k=1}^{T_{hijd}} y_{hijk}, \quad \hat{X}_{hid} = \frac{M_{hi}}{m_{hi}} \sum_{j=1}^{m_{hid}} \sum_{k=1}^{T_{hijd}} x_{hijk}, \quad \hat{X}'_{hid} = \frac{M_{hi}}{m'_{hi}} \sum_{j=1}^{m'_{hid}} \sum_{k=1}^{T_{hijd}} x_{hijk} \; .$$

Contd…

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

$$b_{A2d} = c_{A2xyd} / c_{A2xxd} \, ,$$

Where,

$$c_{A2xyd} = \sum_{h=1}^{H} \left(\frac{1}{n_h} - \frac{1}{n_h{}'} \right) \left(p_{hd}\, s_{bxyhd} + p_{hd}\, q_{hd}\, \hat{X}_{hd} \hat{Y}_{hd} \right),$$

$$c_{A2xxd} = \sum_{h=1}^{H} \left(\frac{1}{n_h} - \frac{1}{n_h{}'} \right) \left(p_{hd}\, s_{bxhd}^{2} + p_{hd}\, q_{hd}\, \hat{X}_{hd}^{2} \right),$$

Contd...

$$p_{hd} = \frac{n_{hd}}{n_h}, \quad q_{hd} = 1 - p_{hd},$$

$$s^2_{bxhd} = \frac{1}{n_{hd}-1}\sum_{i=1}^{n_{hd}}\left(\frac{M_{hi}\, p_{hid}\, \overline{x}_{hid}}{z'_{hi}} - \hat{X}_{hd}\right)^2,$$

$$s_{bxyhd} = \frac{1}{n_{hd}-1}\sum_{i=1}^{n_{hd}}\left(\frac{M_{hi}\, p_{hid}\, \overline{x}_{hid}}{z'_{hi}} - \hat{X}_{hd}\right)\left(\frac{M_{hi}\, p_{hid}\, \overline{y}_{hid}}{z'_{hi}} - \hat{Y}_{hd}\right),$$

$$\overline{x}_{hid} = \frac{1}{m_{hid}}\sum_{j=1}^{m_{hid}} x_{hij.d}, \quad \overline{y}_{hid} = \frac{1}{m_{hid}}\sum_{j}^{m_{hid}} y_{hij.d},$$

$$\hat{X}_{hd} = \frac{1}{n_{hd}}\sum_{i=}^{n_{hd}}\frac{M_{hi}\, p_{hid}\, \overline{x}_{hid}}{z'_{hi}}, \hat{Y}_{hd} = \frac{1}{n_{hd}}\sum_{i=1}^{n_{hd}}\frac{M_{hi}\, p_{hid}\, \overline{y}_{hid}}{z'_{hi}}.$$

Contd…

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

$$\hat{V}\left(\hat{Y}_{lr2d}\right)=\sum_{h=1}^{H}\left(\frac{1}{n_h}-r_{A2d}^2\left(\frac{1}{n_h}-\frac{1}{n_h^{'}}\right)\right)\left(p_{hd}\, s_{byhd}^2+p_{hd}\, q_{hd}\, \hat{Y}_{hd}^2\right), \quad (16)$$

Where,

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

- and c_{A2yyd} is in same functional form as c_{A2xxd} defined earlier
- Then, the estimator of percentage standard error of the proposed regression estimator of total area under d^{th} mixture, $\hat{Y}_{lr2d}$, is given by

Contd…

$$\% \widehat{SE}(\hat{Y}_{lr2d}) = \frac{\sqrt{\hat{V}\left(\hat{Y}_{lr2d}\right)}}{\hat{Y}_{lr2d}} \times 100. \qquad (17)$$

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

$$\hat{Y}_{c2} = \sum_{d=1}^{D^*} \hat{Y}_{lr2d} \qquad (18)$$

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

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

Contd…

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

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

$$\%\,\widehat{SE}(\hat{Y}_{c2})=\frac{\sqrt{\hat{V}\left(\hat{Y}_{c2}\right)}}{\hat{Y}_{c2}}\times 100. \qquad (19)$$

Contd…

Estimation of crop yield under mixed cropping using stratified two stage two phase sampling design framework under area frame approach

- For crop yield estimation, the sample selected for crop area estimation is used as sampling frame for selection of sample for yield estimation.
- In each of the sampled Enumeration Area in a sub-district, first, a list of parcels growing different mixtures is prepared using the sample observed for area enumeration

Contd…

- Suppose, for "Estimation of crop area" by Area Frame Approach under stratified two stage cluster sampling design framework (as discussed earlier), in h^{th} sub-district (Stratum) from the set of N_h Enumeration Areas (PSUs), a sample of n_h Enumeration Areas was selected.
- Then, a sub-sample of segments was selected in each selected Enumeration Area and all the parcels of selected segments of selected tracts were completely enumerated for area estimation under a specific crop mixture.
- While surveying the parcels for area enumeration, a list of parcels growing different crop mixtures was prepared in an Enumeration Area.

Contd…

- Let, there be m_{hid} parcels where the d^{th} crop mixture is grown out of total m_{hi} sampled parcels in the i^{th} EAs of the h^{th} sub-districts,

$$\forall\ i = 1,\ldots, n_{hd};\ \ d = 1,\ldots, D; h = 1,\ldots, H.$$

it is notable that, segment wise crop mixture list need not be prepared

- In the first phase of sampling, from the set of n_h EAs (PSU)
- Let, $n_h^{"}$ = no. of EAs selected by SRSWOR design. Within the selected EA, for each of the d^{th} mixture, ,
- $m_{hid}^{"}$ = no. of parcels (SSUs) selected by SRSWOR from the m_{hid} parcels for recording area

Contd…

- From the selected parcels farmers' eye estimates of harvested yield of the crop under the mixture has to be recorded.
- Let, x_{hijd} = eye-estimated harvested yield of the crop in the j^{th} parcel growing the d^{th} mixture in the i^{th} EA of the h^{th} sub district
- In the second phase of sampling,
- n_h''' = no. of EAs selected by SRSWOR design from the n_h'' selected EA
- and within each selected EA, for each mixture $d = 1, 2, \ldots D$
- let m_{hid}''' = no. of parcels selected from m_{hid}'' parcels in the i^{th} EA of the h^{th} sub-district.
- In the final sample, it can be observed that out of n_h''' sample EAs, n_{hd}''' EAs are following the d^{th} mixture, $\forall d = 1, 2, \ldots D$

Contd...

- Now, from all these m'''_{hid} parcels in the final sample, estimates of the harvested yields of a crop under a specific mixture are obtained by conducting CCE etc. for mixed cropping.
- Let, y_{hijd}=harvested yield of the crop under d^{th} mixture from the j^{th} parcel in the i^{th} EA of the h^{th} sub district.
- By following double sampling approach, utilizing the information from the eye-estimated yields and the harvested yield of the sampled parcels under a specific mixture (domain), a regression estimator of the average crop yield under a specific crop mixture for a district can be given by

Contd…

$$\bar{y}_{2lrd} = \bar{y}_{2d} + b_{2d}(\bar{x}'_{2d} - \bar{x}_{2d}) \quad (20)$$

Where,

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

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

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

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

Contd…

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

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

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

$$s_{bxyhd} = \frac{1}{n_{hd}''' - 1}\sum_{i=1}^{n_{hd}'''}(\bar{x}_{hid} - \bar{x}_{hd})(\bar{y}_{hid} - \bar{y}_{hd}),$$

Contd...

$$s^2_{xhid} = \frac{1}{m'''_{hid} - 1} \sum_{j=1}^{m'''_{hid}} \left(x_{hijd} - \bar{x}_{hid} \right)^2,$$

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

$$p_{hd} = \frac{n'''_{hd}}{n'''_{h}}, \quad q_{hd} = 1 - p_{hd},$$

$$p_{hid} = \frac{m'''_{hid}}{m'''_{hi}}, \quad q_{hid} = 1 - p_{hid},$$

$$\bar{x}_{hid} = \frac{1}{m'''_{hid}} \sum_{j}^{m'''_{hid}} x_{hijd}, \quad \bar{y}_{hid} = \frac{1}{m'''_{hid}} \sum_{j}^{m'''_{hid}} y_{hijd},$$

Contd…

$$\bar{x}_{hd} = \frac{1}{n'''_{hd}} \sum_{i=}^{n'''_{hd}} \bar{x}_{hid}, \qquad \bar{y}_{hd} = \frac{1}{n'''_{hd}} \sum_{i=1}^{n'''_{hd}} \bar{y}_{hid}.$$

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

$$\hat{V}(\bar{y}_{2lrd}) = \frac{1}{H^2} \sum_{h=1}^{H} \left[\left\{ \left(\frac{1}{n'''_h} - \frac{1}{N_h} \right) \left(p_{hd} s^2_{byhd} + p_{hd} q_{hd} \bar{y}^2_{hd} \right) + \frac{1}{n'''_h N_h} \sum_{i=1}^{n'''_{hd}} \left(\frac{1}{m'''_{hid}} - \frac{1}{M_{hid}} \right) s^2_{yhid} \right\} \right.$$
$$\left. - r^2_{2d} \left\{ \left(\frac{1}{n'''_h} - \frac{1}{n''_h} \right) \left(p_{hd} s^2_{byhd} + p_{hd} q_{hd} \bar{y}^2_{hd} \right) + \frac{1}{n'''_h n''_h} \sum_{i=1}^{n'''_{hd}} \left(\frac{1}{m'''_{hid}} - \frac{1}{m''_{hid}} \right) s^2_{yhid} \right\} \right], \qquad (21)$$

Contd…

$$\% \widehat{SE}(\bar{y}_{2lrd}) = \frac{\sqrt{\hat{V}(\bar{y}_{2lrd})}}{\bar{y}_{2lrd}} \times 100. \quad (22)$$

$$s^2_{byhd} = \frac{1}{n'''_{hd} - 1} \sum_{i=1}^{n'''_{hd}} \left(\bar{y}_{hid} - \bar{y}_{hd}\right)^2, \quad s^2_{yhid} = \frac{1}{m'''_{hid} - 1} \sum_{j=1}^{m'''_{hid}} \left(y_{hijd} - \bar{y}_{hid}\right)^2,$$

$$r^2_{2d} = \frac{c^2_{2xyd}}{c_{2xxd}\, c_{2yyd}},$$

and c_{2yyd} is in same functional form as c_{2xxd} as defined earlier

- The estimator of yield for a specific crop *c* based on all the domains at district level along with its approximate estimator of the variance is given by

Contd…

$$\hat{V}\left(\hat{\bar{Y}}_{c2}\right)=\frac{1}{D^{*2}}\sum_{d=1}^{D*}\hat{V}\left(\bar{y}_{2lrd}\right) \qquad (23)$$

$$\hat{\bar{Y}}_{c2}=\frac{1}{D*}\sum_{d=1}^{D*}\bar{y}_{2lrd} \qquad (24)$$

where the sum is over all those domains containing a particular crop *c* in the different mixtures, $d=1,2,\ldots,D*$

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

$$\%\,S\widehat{E}\left(\hat{\bar{Y}}_{c2}\right)=\frac{\sqrt{\hat{V}\left(\hat{\bar{Y}}_{c2}\right)}}{\hat{\bar{Y}}_{c2}}\times 100.$$

Thank you.

Indian Agricultural Statistics Research Institute Library Avenue, Pusa, New Delhi - 110 012.
E-mail id: dr.ucsud@gmail.com
director.iasri@icar.gov.in