F30

USE OF REMOTE SENSING AND SATELLITE IMAGERY IN ESTIMATING CROP PRODUCTION: MALAWI'S EXPERIENCE

Emmanuel J. Mwanaleza
Ministry of Agriculture, Irrigation and Water Development, Statistics Unit, Malawi

DOI: 10.1481/icasVII.2016.f30

ABSTRACT

Malawi through the Ministry of Agriculture, Irrigation and Water Development (MoAIWD) conducted two studies on the use of remote sensing and satellite imagery to estimate crop production. The objective of the studies was to explore a statistically sound methodology for estimating crop production in order to improve the quality and reliability of agricultural statistics to ensure effective policy formulation and strengthen monitoring of Malawi's agriculture sector performance.

The Ministry piloted two methodologies simultaneously. The two studies were successfully implemented and provided useful lessons to the Ministry in terms of producing credible crop production statistics. The two methodologies had the same general technical principles but different statistical methodologies which had strong implications especially in terms of costs and statistical accuracy and validity.

Malawi currently uses a household sample survey known as Agricultural Production Estimates Survey (APES) to estimate agricultural production. APES has three modules namely crops, livestock and fisheries.

Great lessons were learnt from the pilot studies. Remote sensing and satellite imagery provide a statistically sound approach for estimating crop production. It can be used as a stand-alone methodology for estimation crop production or integrated with the household sample survey to develop a more robust methodology for estimating crop production. The use of remote sensing and satellite imagery requires less human resource and period for data collection as compared to the household sample survey. It has also uncovered an opportunity for early objective yield estimation for crops, more precise hactarege estimation, provision of tangible land use distribution criteria to use for the rural road rehabilitation programmimg and generation of crop production maps which can be used for strategic establishment of grain storage facilities in the country.

Based on the results of the pilots and an evaluation of the two pilot methodologies, Malawi has three options for incorporating the use of remote sensing and satellite imagery in its programmes for crop production estimation to improve accuracy of crop production statistics. The first option is complete adoption of the remote sensing and satellite imagery approach for substitution. This involves combining the two methodologies to make one strong methodology and replace it with crop production module of the traditional household survey. The farm household data which is generated through crops module should be collected through the livestock module. The second option is partial adoption of the use of the technology. This involves integrating some modules of the remote sensing and satellite imagery methodologies into the household survey. The last option is to completely adopt the use of remote sensing and satellite imagery for complementarily. This involves using the combined remote sensing and satellite imagery methodology on a regular period to be providing benchmark data for crop estimates.

Key words: crop production estimates, agricultural statistics

1.0 BACKGROUND

The Ministry of Agriculture, Irrigation and Water Development (MoAIWD) annually conducts the annual sample survey of agricultural popularly known as Agricultural Production Estimates Survey (APES). APES also referred to as *list frame methodology* in this paper is a household sample survey conducted to provide important information to the Government of Malawi and its stakeholders for formulating policies and planning of various activities in the agriculture sector. The survey covers three main agricultural sub-sectors, namely: crops, livestock and fisheries. Data collection for the survey is done by extension officers located in various areas across the country.

The survey is conducted in three rounds every year. The first round is conducted from September of the preceding year to January of the current year. Figures from the first round estimates are based on farmers' intentions on crops to be grown and related hectarage. The results from the first round may not conclusively inform the ultimate agricultural production as farmers' intentions can change in the course of implementing respective farm activities. Weather conditions and related parameters may also change in the course of the agricultural season. However, results of the first round provide early warning signals on national food security so that policy makers in the public, private and non state sectors can make sound strategies regarding impending food situation. The first round also involves collection of livestock and fisheries data.

The second round is conducted from February to March and focuses on verification and adjustment of area measurement of crops grown by the sampled agricultural households and results obtained are used to determine crop area for the season.

The third round of the survey which is normally considered as the final round is undertaken during harvesting period from April to May. The third round mainly involves weighing of the harvest to obtain actual yield for crops based on the sampled households. This round includes collection of livestock data as well.

Despite successful implementation of the methodology, it has been observed that there is need to improve certain critical weak areas of the methodology or review the whole methodology to response to the current situation; and adopt and employ innovative and emerging technologies in the production of agricultural statistics. The list frame methodology was designed over twenty years ago. Since then a lot of things have changed. For instance, the number of farm households has more than doubled which requires an adjustment of the sample size. The area under cultivation has also tremendously increased. On the other hand, the number of extension officers responsible for field data collection has greatly decreased due to high labour turn-over.

In a bid to improve crop production statistics, Malawi initiated the use of remote sensing and satellite imagery in the estimation of crop production in 2010. The initial step of the initiative was to conduct a nationwide survey on land cover and use. In 2012 the country piloted a project on crop area in few districts by using point frame approach. The key outcomes of the pilot were that the point frame based methodology proved to be technically feasible for the country and statistically very valuable given the precision of the crop area estimates attained. In 2013 another methodology which used area frame based approach was identified and piloted in few selected areas of the country.

In 2014/15 agricultural season the country simultaneously piloted the two methodologies national wide. The two methodologies had the same general technical principles but different statistical methodologies which had strong implications especially in terms of costs, statistical accuracy and validity. Results of the two pilot methodologies were compared with results from the list frame methodology.

The point frame methodology was developed by a consortium of ITA and EFTAS from Italy and Germany respectively while area frame methodology by Airbus from France. Both methodologies were implemented in collaboration with the Ministry of Agriculture, Irrigation and Water Development

2.0 OBJECTIVES OF THE PILOT PROJECTS

The objective of the studies was to explore a statistically sound methodology for estimating crop production in order to improve the quality and reliability of agricultural statistics to ensure effective policy formulation and strengthen monitoring of Malawi's agriculture sector performance.

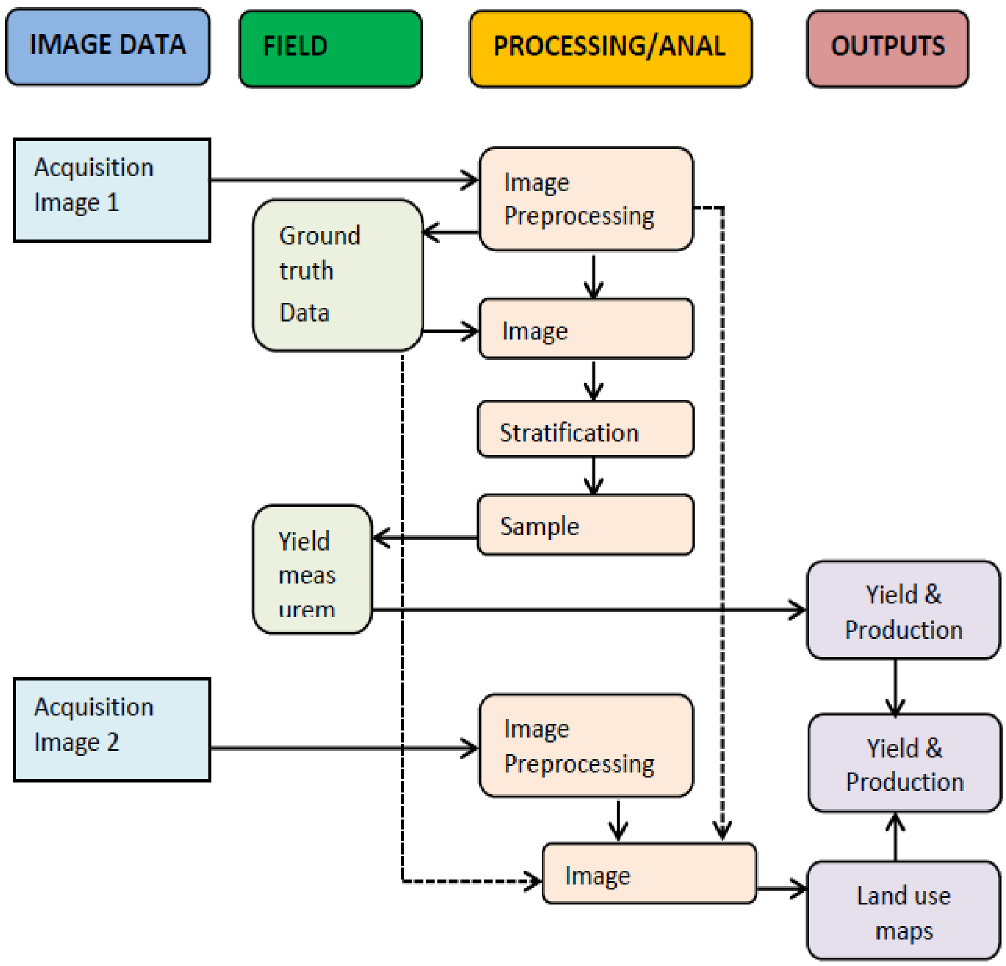
3.0 BRIEF DESCRIPTION OF THE METHODOLOGIES

3.1 Area Frame Methodology

The project used an area frame design. An area frame provides a very efficient tool to survey a territory for various purposes (precise acreage estimation of small crops, yield surveys, production practices, etc.). In the case of Malawi project it was used to estimate maize yields by use of crop cutting. Hectarage estimation was done by the use remote sensing.

Prior to the field data collection, land-use mapping was done using a classification of pairs of satellite images. The classification technique was implemented in three major steps: acquisition of the images, ground truth data collection and classification at the pixel level (Figure 1). To produce the land-use mapping, at least two SPOT 6/7 images at 6m resolution were acquired by district at different times. The first images were acquired from February 19 to April 30, 2015 and the second images were acquired from April 11 and June 9, 2015.

Figure 1: Flow Diagram of the Process



The timeframes for image acquisition were defined with regards to the known local crop calendars. Field surveys were carried out during the first mission in the 15 districts to collect ground truth information required for the calibration of the image classification process and then commonly available Geographical Information System (GIS) and Remote Sensing software (IDRISI, Quantum GIS/GRASS and ArcGIS) were used to perform all data preparation, image integration and classifications. A total of 1551 out of the targeted 1,705 cells were surveyed in all the targeted 27 districts. In each cell two fields were sampled for crop cutting. Data on yield estimate and related information were collected and sent by email to the supervisor on daily basis who after data quality control transmitted the final data to the Ministry headquarters.

Figure 2: GIS Survey Interface Showing Sampled Cells

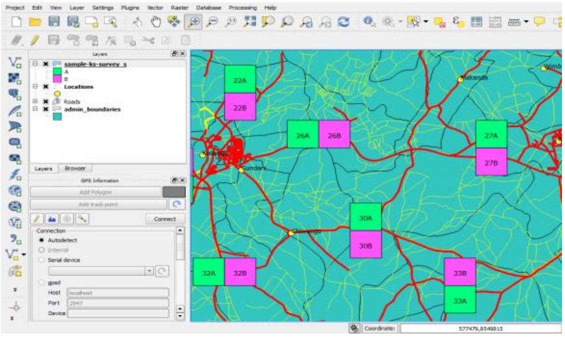
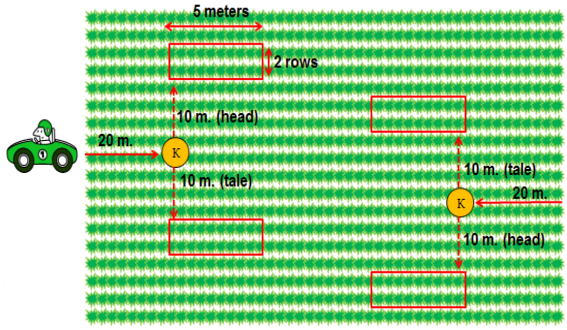


Figure 3: Sample Field with Positions for Crop Cutting



3.2 Point Frame Methodology

The methodology is point frame based whereby the surveyors were required to visit the sampled points, observe within a 3m radius the crop grown at that particular point and record the data. A total of 24000 points were sampled for the survey. Sampled points were classified into clusters of 16 points each with a distance of 250m from one point to the other. For yield estimation, farmer interviews and crop cutting were conducted to assess productivity.

Figure 4: Sample of a Cluster

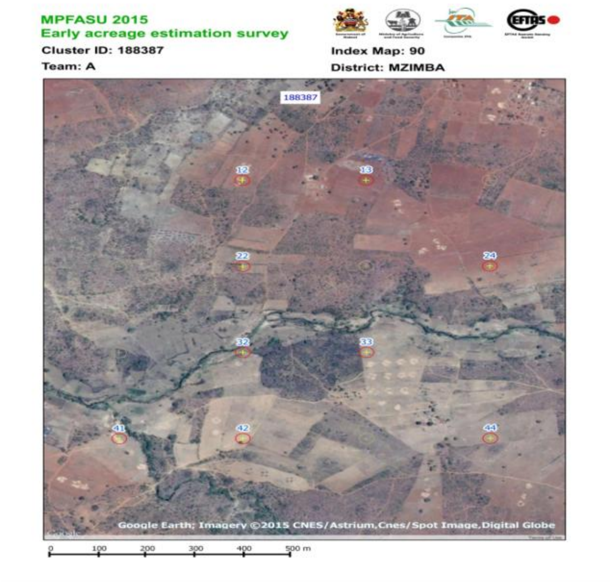
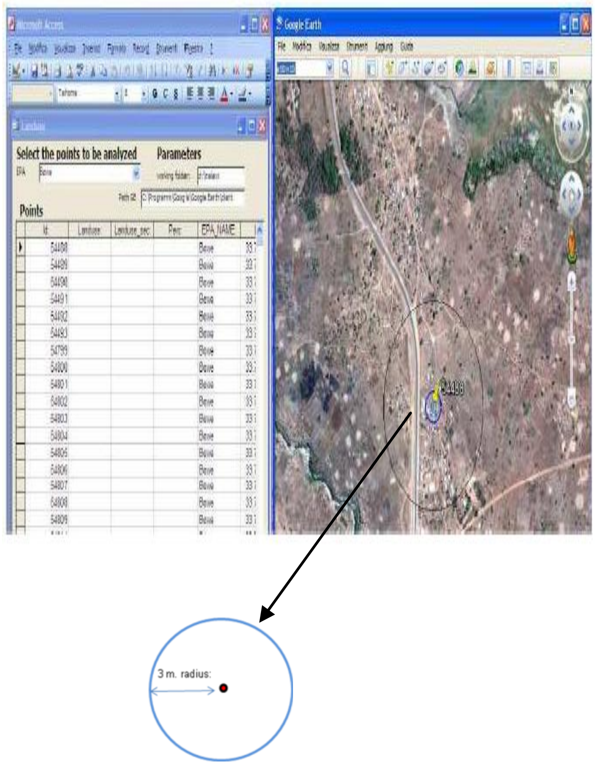


Figure 5: Sample of a Point



NOTE: a point is a geo-location of the area surveyed. The crop is observed within 3 meters of the point. With the guidance of the Global Positioning System (GPS), upon reaching the point the enumerator observes and records on the type of crop within the radius of the point. For mixed cropping the enumerator has to judge the percentage of each crop in the point

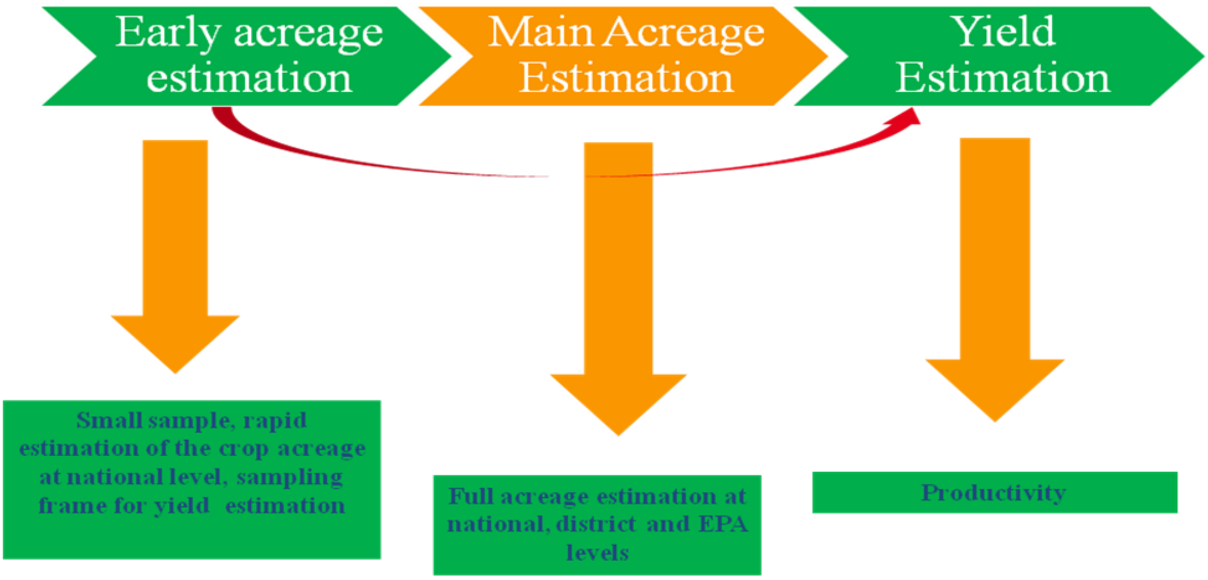
The project had three main modules:

Module1: **Early Crop Acreage Estimation:** the objective of the module was to provide a forecast of crop area at national level. The module also provided a sampling frame for selection of points for yield estimation using crop cutting method**.** The module objectives was achieved by using a sample of 300 clusters on the agricultural land with 16 sampling units (points) inside a square grid spaced 250 x 250 m.

Module 2: **Full Crop Acreage Estimation**: the second project component was designed in order to obtain crop acreage estimation across the country and for the crops of the country with a Coefficient of Variation (CV) of 4-5% for maize. The ground survey was planned to be carried out on 25,000-27,000 sampling units (points) grouped in 1500 clusters and randomly allocated separately across the country.

Module 3: **Yield Estimation**: the objective of this module was attained by interviewing 1,200 farmers located within the 300 clusters. The farm interviews collected among others data on quantity the farmer had harvest from the garden, socio-economic information and sizes of the garden. In addition crop cutting was done in selected 20 gardens within the same clusters to complement farm interviews data.

Figure 6: Main Modules of the Project



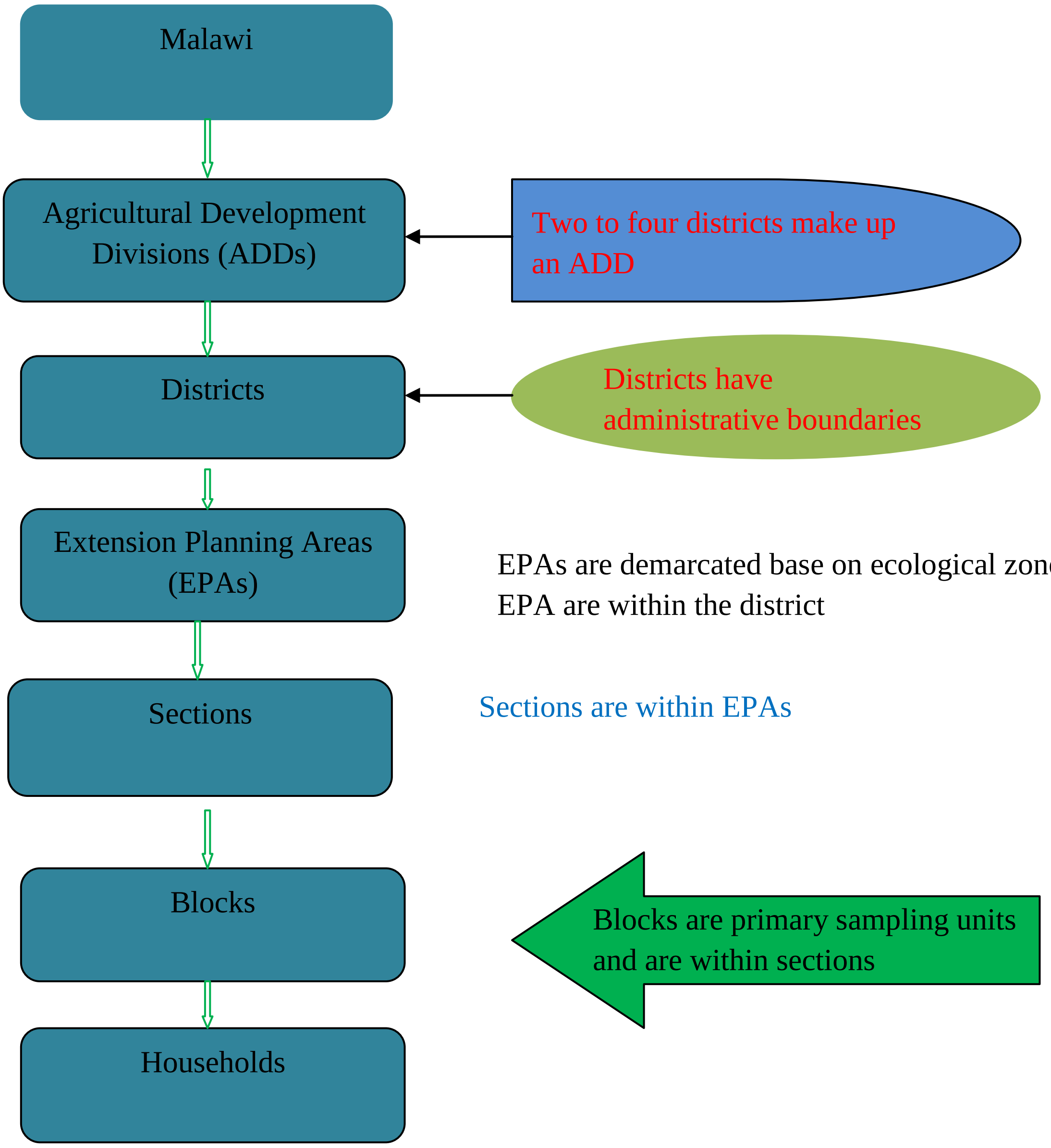
Field data collection was done by forty surveyors for a period of four weeks. Data collection and data entry was done at the same time using tablets. Data was submitted to a cloud archive server and centrally downloaded for monitoring of data flow and quality control.

3.3 List Frame Methodology

The survey begins in September every year up to April in the preceding year. The survey uses a stratified two-stage systematic sampling. Extension Planning Areas (EPAs) making up a district constitute the strata. Each EPA consists of Sections and individual Sections comprise Blocks. The Block is the Primary Sampling Unit (PSU) used in the survey (Figure 6).

The first step in sample identification involves the selection of 25 percent of the Blocks in each EPA using a systematic random method. The second step requires the identification of a Secondary Sampling Unit (SSU) of agricultural households drawn from the selected Blocks. In each selected Block, all households are listed. From the list, agricultural households are serially listed and a survey sample of 20 percent, but in any case not more than 15 households, is identified again using systematic random sampling.

Figure 7: Sampling Structure



Extension officers, who are data collectors for the survey, are required to list all households in the selected Blocks and identify agricultural households. The officer indicates if the household grows field and/or horticultural crops, keeps livestock and do any fish farming activities. For crop module, extension officers are required to conduct garden measurement exercise using the GPS in the sampled households.

For yield assessment, the estimation for the first and second rounds are subjective based on extension officer's judgment on crop development. Based on the crop stand, the extension officer is required to make a judgment whether the yield will remain the same, go up or down and by what percentage. In the third round the actual yield is estimated using crop cutting. Livestock data is collected by actual counting of the livestock from the household. Fisheries data is collected from both fish farming/aquaculture from the sampled households and capture fisheries from natural water bodies. Data processing and analysis is done at the EPA level, and then data aggregation processes are done at the district, Agricultural Development Divisions (ADDs) and national levels.

4.0 RESULTS

Results of the two projects compared with the results from the list frame approach are shown in Tables 1,2 and 3. The comparisons are made for maize only because area frame methodology focused on maize crop only. Maize is the major staple food in Malawi. The comparisons are from rainfed production only for both results as the pilot projects did not take into account winter cropping as done in the list frame methodology.

Table 1 provides results on hectarage estimation from all the three methodologies. From point frame methodology, hectarage was estimated at 1,396,477 hectares lower by 7 percent as compared to 1,499,340 hectares from the list frame. The area frame estimated hectarage at 1,812,908 hactares giving a difference of 21 percent as compared to the list fame approach. Generally, results from point frame estimation are lower across the ADDs while figures from area frame are higher as compared to the results from list frame methodology. Comparisons further show that results are close at national level with great variations at ADDs.

Table 1: Results of the hectarage estimation

		List frame Vs Point frame			List frame Vs Area frame			
ADD	List frame	Point frame	Diff	% Diff	List frame	Area frame	Diff	% Diff
Blantyre	260,999	197,610	63,389	24	260,999	266,646	-5,647	-2
Kasungu	341,783	330,423	11,360	3	341,783	402,368	-60,585	-18
Karonga	52,962	30,843	22,119	42	52,962	57,481	-4,519	-9
Lilongwe	334,212	397,450	-63,238	-19	334,212	473,174	-138,962	-42
Machinga	266,301	226,538	39,763	15	266,301	322,229	-55,928	-21
Mzuzu	144,210	145,200	-990	-1	144,210	172,340	-28,130	-20
Salima	61,588	50,486	11,102	18	61,588	59,665	1,923	3
Shire Valley	37,285	17,928	19,357	52	37,285	59,005	-21,720	-58
National	1,499,340	1,396,477	102,863	7	1,499,340	1,812,908	-313,568	-21

In terms of productivity, the point frame estimated a national maize yield of 1430 kgs lower by 129kgs representing 5 percent difference as compared to the list frame. The area frame estimated national maize yield of 1470kgs versas 1508kgs from list frame, giving a difference of 3 percent (Table 2). In both cases national maize yield from list frame is slightly higher as compared to the two pilot methodologies.

Table 2: Results of the Yield Estimation

	List frame Vs Point frame				List frame Vs Area frame			
ADD	List frame	Point frame	Diff	% Diff	List frame	Area frame	Diff	% Diff
Blantyre	1,484	612	1,163	66	1,484	1,512	263	15
Kasungu	1,790	1,543	247	14	1,790	1,825	-35	-2
Karonga	2,261	1,586	675	30	2,261	2,228	33	1
Lilongwe	1,748	1,975	-227	-13	1,748	1,259	489	28
Machinga	774	1,160	-386	-50	774	1,050	-276	-36
Mzuzu	1,452	1,014	438	30	1,452	1,676	-224	-15
Salima	1,828	1,137	691	38	1,828	1,770	58	3
Shire Valley	816	446	370	45	816	1,210	-394	-48
National	1,508	1,430	129	5	1,508	1,470	89	3

Results on national maize production are presented in Table 3. The point frame methodology estimated national maize production at 1,996,428 metric tons while the list frame estimated 2,261,581 metric tons. The area frame methodology estimated the production at 2,665,492. The results show that figures are close at national level with great variations at lower levels. Possible reasons for such differences are small sample sizes for the pilot methodologies, hence, there is need to increase the sample size to achieve good precision at lower levels. However, an investigation is required to establish factual causes for the differences. As the case with hacterage figures, generally results from point frame estimation are lower across the ADDs while figures from area frame are higher as compared to results from the list frame.

Table 3: Results of the Production Estimation

	List frame Vs Point frame				List frame Vs Area frame			
ADD	List frame	Point frame	Diff	% Diff	List frame	Area frame	Diff	% Diff
Blantyre	387,239	136,639	326,536	70	387,239	403,045	60,130	13
Kasungu	611,955	498,230	113,725	19	611,955	734,314	-122,359	-20
Karonga	119,772	46,608	73,164	61	119,772	128,043	-8,271	-7
Lilongwe	584,066	778,083	-194,017	-33	584,066	595,830	-11,764	-2
Machinga	206,120	270,997	-64,877	-31	206,120	338,408	-132,288	-64
Mzuzu	209,432	178,067	31,365	15	209,432	288,823	-79,391	-38
Salima	112,565	76,210	36,355	32	112,565	105,606	6,959	6
Shire Valley	30,432	11,593	18,839	62	30,432	71,423	-40,991	-135
National	2,261,581	1,996,428	341,089	12	2,261,581	2,665,492	-327,975	-18

5.0 LESSONS LEARNT AND RECOMMENDATIONS

A number of important lessons were learnt from the pilot studies. Remote sensing and satellite imagery provide a statistically sound approach for estimating crop production. It can be used as a stand-alone methodology for estimation crop production or integrated with the household sample survey to develop a more robust methodology for estimating crop production. The use of remote sensing and satellite imagery requires less human resources and period for data collection as compared to the household sample survey.

From the results of the pilots and lessons learnt from the pilot methodologies, Malawi had three options for incorporating the use of remote sensing and satellite imagery in its programmes for crop production estimation to improve accuracy of crop production statistics. The first option was complete adoption of the remote sensing and satellite imagery approach for substitution. This involves combining the two methodologies to make one strong methodology and replace it with crop production module of the traditional household survey. The farm household data which is generated through crops module should be collected through the livestock module. The second option was partial adoption of the use of the technology. This involves integrating some modules of the remote sensing and satellite imagery methodologies into the household survey. The last option is to completely adopt the use of remote sensing and satellite imagery for complementarily. This involves using the combined remote sensing and satellite imagery methodology on a regular period to be providing benchmark data for crop estimates.

Based on the evaluation of the two methodologies Malawi opted for partial adoption of the use of the technology by integrating some modules of the point and area frame methodologies into the list frame methodology. To maintain livestock and fisheries components of the list frame methodology, there is need to go for multi-frame approach. This means livestock and fisheries data should be collected from the list frame and crop production data from combined point frame and list frame methodologies.

The evaluation also recommended adoption of number of lessons learnt from the projects aiming at improving agricultural production statistics. The recommendations were categorized into short and long terms.

5.1 Short Term Recommendations

i) Use of Information, Communication and Technology (ICT) in Data Collection and Transmission: There is need to introduce electronic data transmission system for agricultura surveys. This is one of the important lessons learnt from both methodologies. Electronic data acquisition and transmission has proven to promote quick and error free data submission. There is need to use the existing central server located in the Ministry headquarters for data submission. Daily data transmission by data collectors should be made possible by distributing tablets (i.e., with GPS and internet access) to each of the data collector

ii) Development of electronic database: There is need to develop an electronic database for agricultural production survey. All data should be centrally stored and managed. By having micro-data, the Ministry will be able to derive estimates with coefficient of variations to determine precision of estimates for each crop at all levels. The database should be programmed so that estimates are calculated automatically as data is being transmitted to the database. The data collection program should be designed in such a way that errors are detected automatically as data is being transmitted hence reducing period for data cleaning.

iii) Introduction of moisture content measurement for crops: The current methodology does not measure moisture content for crops. It is recommended that weighing of crops should be done when the crop is completely dry. However, it is not certain whether this protocol is strictly followed by the extension officers. If it is not followed, this gives a chance for overestimation. Even if it is followed, this means that the extension officers have to wait for the crop to dry in cases of early harvesting, hence, delay in submission of the data. It is strongly recommended that moisture content measurement as learnt from the area frame methodology be introduced to improve data quality and facilitate early submission of the data.

5.2 Long-Term Recommendations

i) Use of point frame approach for hectarage estimation: The point frame methodology should be adopted for hectarage estimation. However, to achieve good estimates at all levels, there is need to increase the sample size for points. The point frame hectarage estimation has proven to be more accurate, quick, easy, less time consuming and requires less human resource.

ii) aintain yield estimation method from the list frame: Crop cutting method from list frame methodology has proven to produce credible results. The method is applicable to all crops. Yield estimation methods from the pilot methodologies either require some adjustments (point frame methodology) or separate methods for non-cereal crops (area frame methodology).

iii) Revision of sampling frame for the list frame: The current methodology samples 15 households in each block. This means that all blocks are treated as homogeneous entities. This may result in oversampling in areas where there are less agricultural households and under sampling in areas with more households. The sampling plan should be revised. The sampling plan should be based on agricultural land use, meaning, a sample of households should be allocated according to the agricultural land in that particular area. Stratification taking advantage of differences in agricultural land use intensity will provide a significant increase in precision. Therefore, from the area frame methodology, revising the sampling methodology under list frame based on actual agricultural land use density rather than population estimates is recommended.

iv) Development of strong capacity building programme: If the recommendations are approved, there will be need for strong capacity building in the Ministry mainly through short and medium-term training. Arrangements should be made with the Development Partners such as Food and Agricultural organization (FAO) of the United Nations for local technical support in sampling, GIS and image analysis.

6.0 CONCLUSION

Great lessons have been learnt from the pilot studies which can improve crop area and production statistics. Application of remote sensing and satellite imagery technologies in crop production estimation has proven to provide quick, reliable and timely statistics. It has also uncovered an opportunity for early objective yield estimation for crops, more precise hactarege estimation, provision of tangible land use distribution criteria to use for the rural road rehabilitation programmimg and generation of crop production maps which can be used for strategic establishment of grain storage facilities in the country. However, to ensure successful implementation of the lessons leant, there is great need for strong capacity building programme in the area of GIS and remote sensing.