Role of Earth Observations for Crop Area Estimates in Africa. Experiences from the AGRICAB Project

Michele Downie
Consorzio-ITA
Via Farnesi 25/A 55100 Lucca, Italy
studiomappe@gmail.com

David Remotti
Consorzio-ITA
Via Farnesi 25/A 55100 Lucca, Italy
david@agros.it

Tomaso Ceccarelli *
Alterra, Wageningen UR
Droevendaalsesteeg, 6708 PB
Wageningen, the Netherlands
tomaso.ceccarelli@wur.nl

DOI: 10.1481/icasVII.2016.g44b

ABSTRACT

Despite major technological advancements, Earth Observations (EO) are nowadays seldom used in national agricultural statistical systems at an operational level. This is with the exception of a few countries and is especially true in Africa (but for Morocco, South Africa and a few other cases). Moving from small tests confined to research to operational contexts, poses a number of challenges. There are many aspects that still restrict the use of this technology: the high cost of some of the images, the level of expertise, overall organisation and institutional arrangements needed to operationally run the different components of the system.The FP7 AGRICAB project provided an important opportunity to evaluate the role of EO specifically for crop area estimates using spatial frames in Africa. This is with reference to: 1) the construction and stratification of a (point) sampling frame, 2) the support to the ground surveys, and 3) the use of the image classifications to further improve the accuracyof the estimates from the surveys alone. The approach was tested in selected areas in Mozambique, Senegal and Kenya. The contribution of EO for the last aspect cited was evaluated only for the latter two areas, where high geometric resolution images (RapidEye) were used. Sampling frames were built based on high geometric resolution images accessible through Google Earth. This contributed significatively to the frame construction and to increased sampling efficiency through land cover stratification. As to the use of EO for survey preparation/execution, the images also provided invaluable support to the geolocation of sampling units especially if combined with GPS. Finally, EO data were used for reducing the variance of ground estimates. If a high correlation between the classification of the image and the ground truth exists, it is possible to produce estimates with a lower sampling error. From a theoretical point of view a perfect correlation between spectral signatures of a generic crop and the corresponding parcel, would promote pure remote sensing approaches such as pixel counting and sampling frames combined withground surveys unnecessary. Unfortunately this high correlation rarely exists and in most cases satellite images cannot be used directly. Confusion matrixes generated from the supervised classifications of the test areas showed a very low overall accuracy and hence, a very low contribution to the reduction of the variance of the estimates which was evaluated further in terms of net efficiency and cost efficiency. This was to be expected, considering the characteristics of most Sub-Saharan agricultural landscapes, i.e.: small fields vs. pixel size, continuous and mixed cropping, low planting densities. The use of multi-temporal images and the combination of optical and radar EO data can probably increase such correlation although a rigorous cost/benefits analysis would be needed to evaluate its added value also in view of new satellite products (e.g. those provided by the Sentinel missions).

Keywords: spatial frames, satellites, relative efficiency

1. Introduction

There are two main methods to derive crop area statistics from Earth Observations (EO) data: pure remote sensing approaches such as pixel counting, and methods combining field survey data and image classification results (FAO 2015). Pixel counting is the more direct way, although it is often criticized for the bias which can be introduced (Gallego at al., 2008 and Gallego et al., 2010). This seems to be especially true with reference to smallholders farming systems in Sub-Saharan Africa.

Despite major technological advancements, EO are nowadays seldom used in national agricultural statistical systems in an operational way. There are a limited number of countries where this occurs worldwide and, apart from Morocco, South Africa and a few other cases, this is especially true for Africa.

Moving from small tests confined to research to operational contexts, poses a number of challenges. There are many aspects that still restrict the use of this technology: depending on their use, the high cost of some of the images, the level of expertise required, overall organisation and institutional arrangements needed to run operationally the different components of the system.

The FP7 AGRICAB project provided an important opportunity to evaluatethe contribution of EO specifically for crop acreage estimates using spatial frames in Africa. This is with reference to three levels in the methodology: 1) construction and stratification of a point sampling frame, 2) support to the ground survey, and 3) its use to further improve the accuracy of the estimates from the survey alone.

The project initially targeted selected areas in Senegal, Kenya and Mozambique and together with national partners mandated with agricultural statistics in the countries. EO products were effectively used in all areas. However only in the case of Senegal and Kenya, where images with high geometric resolution were available, it was possible to cover all cited levels of application.Due to the cloud cover, for MozambiqueEO products were used only atthe first two levels.

2. Materials and Methods

2.1 Materials

The areas selected were the District (Département) of Nioro Du Rip in Senegal, and the County of Kakamega-Butere in Kenya. As mentioned, only for these areas it was possible to acquire images with high geometric resolution.In Mozambique the area selected was the District (Distrito) of Inharrime. The three areas are indicated in Figure 1.

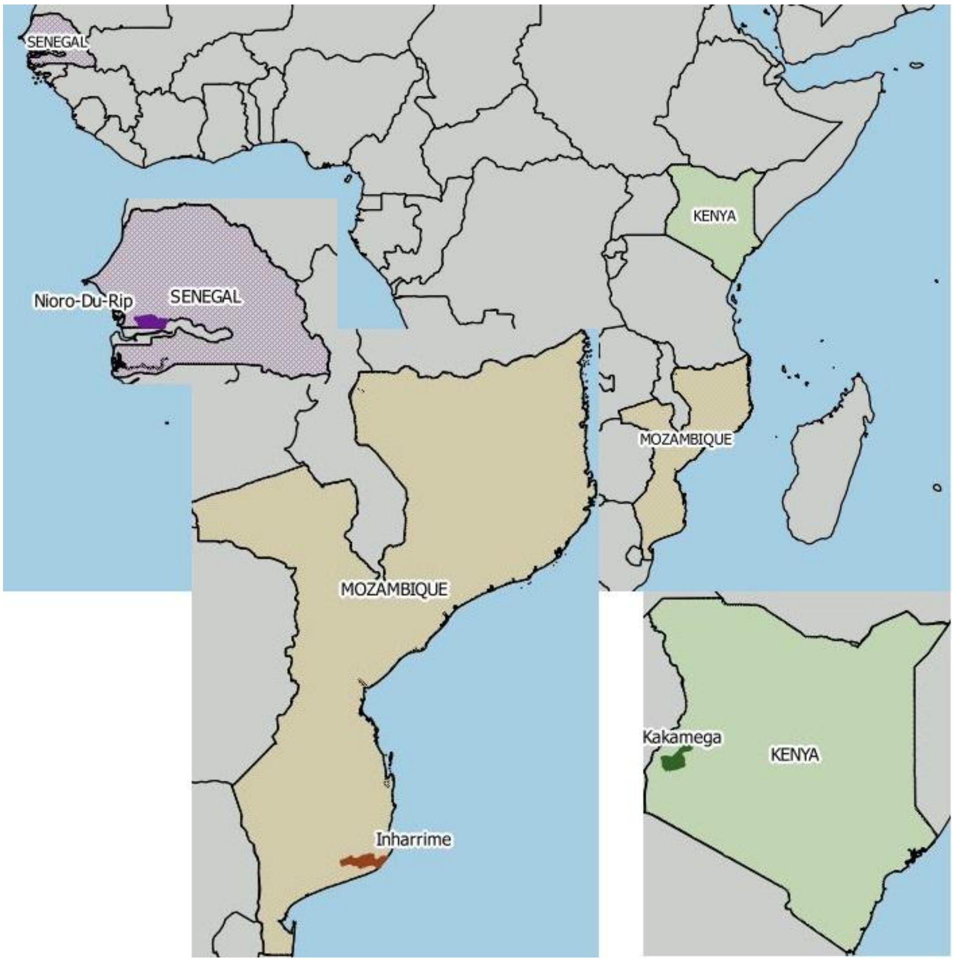

Figure 1: *The three areas of interest*

The EO images used belong to the RapidEye satellite constellation. For the cited study areas the provider programmed the image acquisition at a convenient time with reference to the presence of the targeted crops. Harvesting periods relating to the two area tests were indicated by experts from the respective Ministries of Agriculture who have knowledge of the climate and the characteristics of the selected test areas. In Kakamega-Butere the period was 15 May - 30 June while in Nioro Du Rip, 15 August - 30 September.

Table 1: *Main features of RapidEye images.*

MISSION CHARACTERISTIC	INFORMATION	
Number of Satellites	5	
Spacecraft Lifetime Over	7 years	
Orbit Altitude	630 km in Sun-synchronous orbit	
Equator Crossing Time	11:00 am local time (approximately)	
Sensor Type	Multi-spectral push broom imager	
Spectral Bands	Capable of capturing all of the following spectral bands:	
	Band Name	**Spectral Range (nm)**
	Blue	440 510
	Green	520 590
	Red	630 685
	Red Edge	690 730
	NIR	760 850
Ground sampling distance (nadir)	6.5 m	
Pixel size (orthorectified)	5 m	
Swath Width	77 km	
On board data storage	Up to 1500 km of image data per orbit	
Revisit time	Daily (off-nadir) / 5.5 days (at nadir)	
Image capture capacity	4 million sq km/day	
Camera Dynamic Range	12 bit	

As it discussed in the next sections, satellite images are complementing survey data. A description of the content of such information as well as the way in wich ground data are collected is also given in the next setions.

2.2 Sampling Frame Construction

The sampling frame is the most important element in methodology developed. Its functions are:

- to enumerate all the units of the population;
- to label and stratify the same units based on a limited number of land cover classes;
- to allow the extraction of sampling units for a specific statistical survey;
- to subsequently extrapolate to the universe the values derived from the sample.

The construction of the frame requires the following steps:

Defining the units of the population. These units are represented by geographic points located at the vertices of a grid of 500 x 500m. All points within an administrative boundary represent the population whose parameters (crop area in this case) we intend to estimate.

Building the frame. This implies assigning (labelling) the land cover type to each unit based on very few and simple classes. This is done through on screen visual interpretation of satellite imagery. The images do not necessarily need to be of the same year of the survey, nor need to be taken during the growing season; the higher the resolution, the better, since it allows a more accurate identification of the named classes. In this respect very high resolution (VHR), i.e. sub-metric satellite images freely available on Google Earth or similar sources were found to be the most suitable basis to derive the land cover types related to the points making the population.Only in a few cases high resolution images were not available.

Stratification. The land cover classes assigned to the points in the frame provide a basis for its further stratification.

Satellite images are importantfor the construction of the frame. Without their contribution the said frame would rely on ortho-photographs in digital format (which are simply not available in most African contexts) or be built using only topographic maps. These maps in most cases do not contain information useful for the stratification hence reducing the efficiency of the statistical system as a whole.

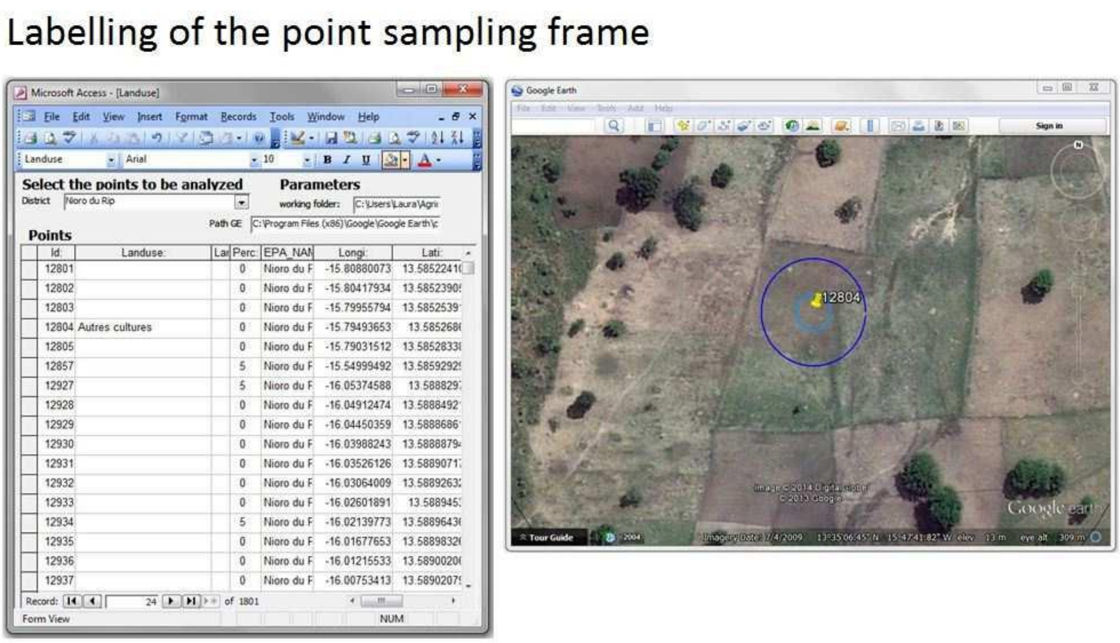

Figure 2: *Labelling of the point*

2.3 Support to ground-survey

The ground survey consists of the identification of crop types and other additional information in each sampling point. In this way objective, verifiable and repeatable data on crop occurrence can be collected based on rigorous sampling schemes. The ground observations are carried out by the surveyor whit reference to an area (usually a circle of fixed radius) around each point.

When carrying out the ground survey, the surveyor is faced with a number of challenges such as poor road network and lack of topographic information. Areas are also often largelyuninhabited,with long distances between points, and may present poor accessibility as well as specific dangers. Therefore the survey must be carefully planned beforehand and constantly revised on the ground.

Recent, very high resolution EO products are required for the purpose.There are usually three types of maps available to the surveyors, all having satellite information as a backdrop image. Examples are given in Figure 3. An “indexmap” is used as a topographic base map to locate the tiles. A “tile map” is a portion of the base map, generally at a 1:50.000 scale, and contains the points which should be visited by a surveyor. A “(sample frame) point map”,generally at a 1:2.000 scale, which is used together with a hand held GPS to identify and access the sampling point. Such map with the sampling units superimposed on satellite images can greatly facilitate the proper recognition of their position. For the geolocation of the points one should not rely exclusively on most GPS used in these types of surveys, which have known limitations in their geometric accuracy.

It is important to underline that the type of land cover and crop type information which needs to be associated to each point, cannot be derived directly from the satellite images used in the construction of the frame and in its stratification. This is due to the fact that such images are usually taken before the ground survey (i.e. when the cloud cover is minimal). Therefore they usually do not allow for a proper identification of the crop type or, if this is the case, may represent crops which have changed over the different cropping seasons.

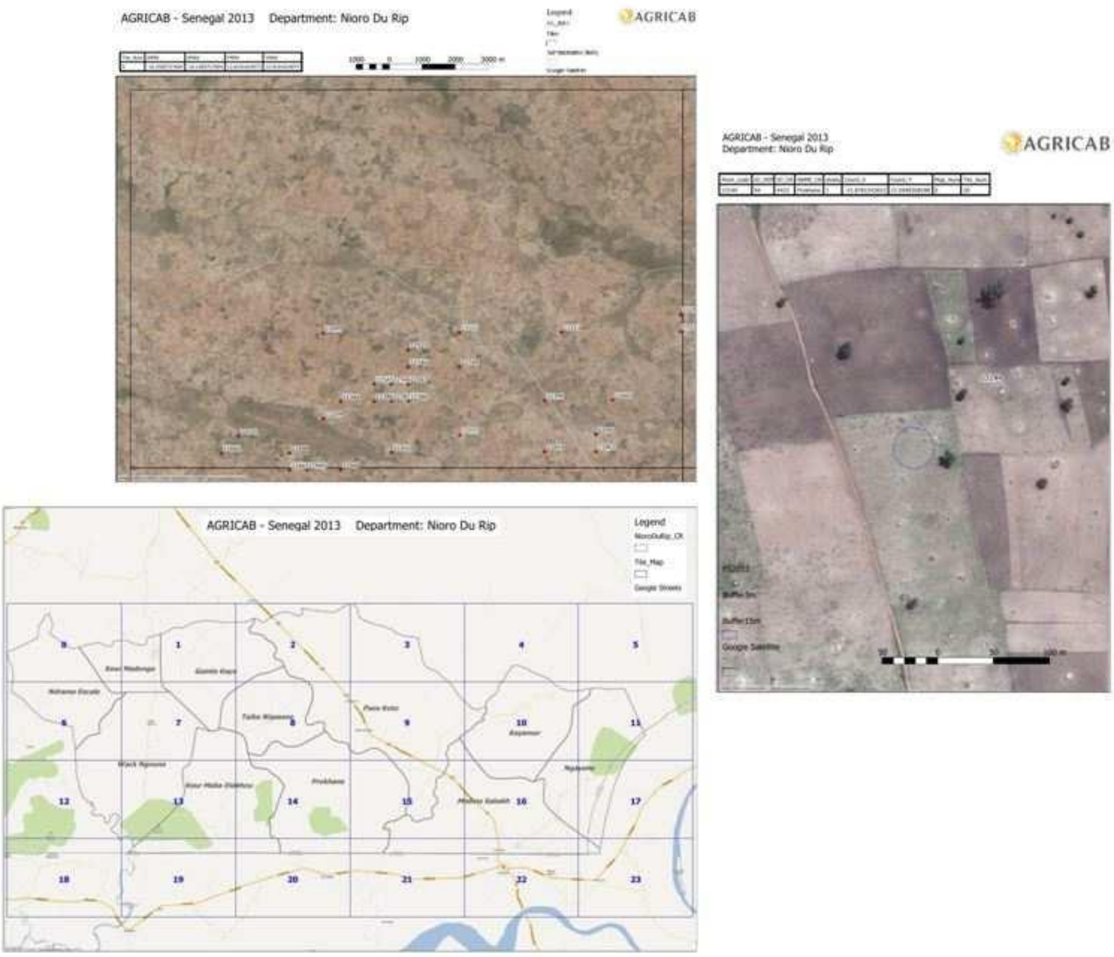

Figure 3: *Field Maps (from bottom left: Index, Tile and Point Map)*

Using the point maps, the surveyor can also plan the best itinerary to reach the different sampling units. Each surveyor has access to a series of field tools (the same field maps, field forms and a GPS). These would allow him a) to reach, as quickly as possible, each sampling unit and to identify with high accuracy (from a geographical point of view)the point of observation of the same unit, b) to identifythe type of crop according to a pre-defined legend and classification rules. An illustration of the field tools is given in figure 4.

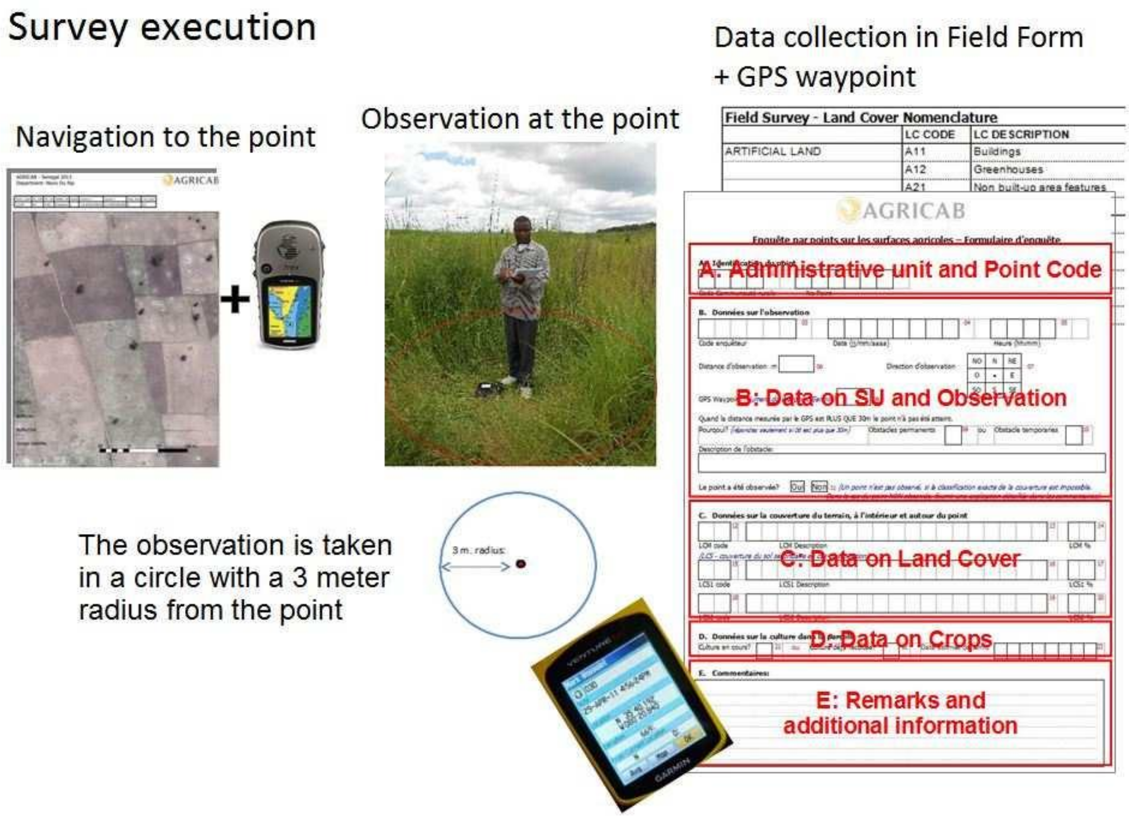

Figure 4: *Survey field tools*

Based on the experience in the different surveys conducted we can conclude that VHR images, combined with other baseline information and the location of the same sampling units provide an invaluable support for survey planning and execution. They make it possible to carry out the survey quickly and rigorously without the need of contacting the farmers.

2.4 Use of the image classifications to further improve the accuracy of the estimates from the ground surveys

This is, at least in principle, one of the most important aspects related to the application of EO data for crop area estimates. If a high correlation between the classified images and the ground truth exists, it is possible to generate crop area estimates characterized by a lower sampling error (expressed in terms of coefficient of variation, or CV). The RapidEye images acquired for the two study areas have been processed with the aim to verify the degree of the correlation between the spectral signatures of the main crop types and the corresponding ground truth observed during the survey. The following procedures have been utilized for processing the satellite images:

Image pre-processing.includes geo-referencing through registration of images, mosaicing of the RapidEye tiles with the same dates and same atmospheric conditions, production of masks to remove non-agricultural areas from the classification (otherwise introducing a bias).

Spectral signature extraction. The final outcomes of this step area number of Regions Of Interest (ROI) which are then used to guide the classification and to provide the ground truth locations needed for generating the confusion matrix.

The extraction was carried out using the observations collected during the surveys. Data varyaccording to the study areas and the respective methodologies. For Kenya the ground truth was based on the "Field Sample Points" grid(around 1.000 points) and on an additional sample of "Photo-clusters"(each cluster is a meshof 100 points with 50 meters spacing). This sample was classifiedby visual interpretation of aerial ortho-photos collected by the Department of Resource Surveys and Remote Sensing (DRSRS) during the survey period. In Senegal the available ground truth came from the "Field Sample Points" grid data (again around 1.000 points collected in the field) and the "Parcel Segments"data collected by the “Direction de l'Analyse, de la Prevision et des StatistiquesAgricoles” (DAPSA) of the Ministry of Agriculture, with the support of the “Centre de SuiviEcologique” (CSE). Crop sample areaswere collectedby GPS during the “Enquêteagricole 2013” in conjunction with the AGRICAB point frame survey. The procedure needs to be adjusted to each area due to differences in the type of the ground truth data i.e. to differences in the classification rules applied and tothe geometry of the data (points or areas). Such information is often collected in difficult environmental conditions and this goes sometimes to the detriment of its quality and accuracy. Moreover important factors related to the spectral signature (phenological state, health conditions, farmingpractices, water stress, etc.) could not always be captured during the surveys.

Supervised classification of the images. The final mask, the mosaics and the ROI dataset were imported in a GIS environment. In Kenya the classification was carried out based on the QGIS and GRASS open source solutions, while in Senegal the software ERDAS was used.The ROI dataset was used to generate the signature files, one for each mosaic. The classification was performed applying the maximum likelihood classifier (pixel based).

Considering the available information it was decided to classify the images focusing on the most representative crops: Sugarcane and Maize for Kenya and Millet, Maize, Sorghum and Groundnuts for Senegal. All the other land use classes were grouped as "other land cover/land use” (LC/LU).

The following tasks were performed for the classification:

- **use ofsub-classes** in order to cover as much as possible the various spectral signatures that characterize each LC/LU class;
- **adding the most important natural/non-agriculture sub-classes** to reduce the variance in the classification of the agricultural classes;
- **aggregating the sub-classes** corresponding to the various spectral classes obtained from the images in LC/LU classes and all the non-agricultural sub-classes as "Other LC/LU";
- **filtering the classified pixels** with a confidence level of at least 66% to70%;
- **calculating the confusion matrix** and the accuracies;

An example of the original image and of the classification is given in Figure 5.

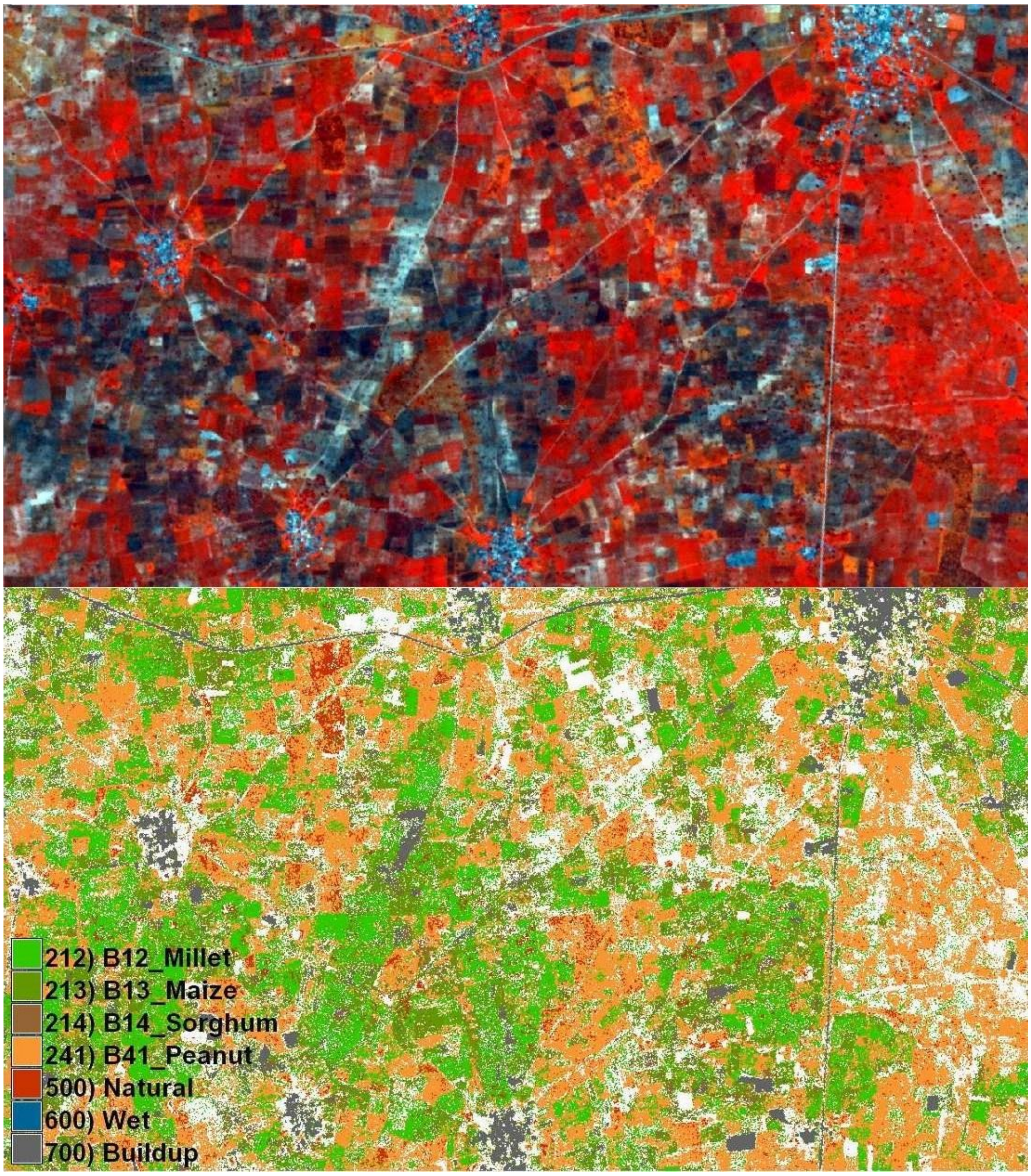

Figure 5: *Example of the original image and the classification in Senegal, Nioro Du Rip. From top: composite RGB 543 and classified raster image*

The accuracies derived from the confusion matrix should not be regarded as a post classification assessment, but rather as an indicator of the validity of the classificationin further reducing the estimates based on ground survey data alone.

Second stratification. The classification provides information on land cover and crop types that could be used to achieve better estimates of crop areas. The principle is that this information, known for the whole of the population and not only for the sampling units, can be used to build a stratification of the population in homogeneous strata where the variance of the estimates is expected to be significantly lower than in the population as a whole.

A new stratification is performed on top of the initial one (see 2.2), using a binary classification for each point in the sample frame.For instance in the case of maize four new strata are derived as a combination of the original strata, and the points classified as maize (or "not-maize") in the image.

3. Results

As shown in the tables 2 and 3 below, the confusion matrices generated in the two different study areas show a very low overall accuracy in the classification: around 45% for Senegal and around 40% for Kenya. For the main crops occurring in each area omission and commission errors are relatively better.

Table 2: *Confusion matrix Nioro Du Rip, Senegal*

		GROUND TRUTH						
	CLASS	B12	B13	B14	B41	NAT	Row Sum	COMMISSION ERROR / User's Accuracy / inclusion
CLASSIFICATION	B12 millet	**56**	18	2	44	7	**127**	55.91%
	B13 maize	40	**17**		28	7	**92**	81.52%
	B14 sorghum	1	1		4		**6**	100.00%
	B41 groundnut	19	13		**92**	8	**132**	30.30%
	NAT	3	1		11	**3**	**18**	83.33%
	Column Sum	**119**	**50**	**2**	**179**	**25**	**375**	
	OMISSION ERROR / Producer's Accuracy / exclusion	52.94%	66.00%	100.00%	48.60%	88.00%	ACCURACY OVERALL	**44.80%**

Table 3: *Confusion matrix Kagamega-Butere (Kenya)*

		GROUND TRUTH						
	CLASS	B13	B35	Z99	Row Sum			
CLASSIFICATION	B13 maize	**10**	11	22	**43**	76.74%	COMMISSION ERROR	User's Accuracy inclusion
	B35 sugar cane	10	**22**	14	**46**	52.17%		
	Z99 Other LC/LU	6	5	**14**	**25**	44.00%		
	Column Sum	**26**	**38**	**50**	**114**			
		61.54%	42.11%	72.00%	OVERALL ACCURACY	**40.35%**		
		OMISSION ERROR Producer's Accuracy exclusion						

Possible reasons behind these low accuracies can be ascribed to several factors. Some factors are of general applicability while others are especially important in the agricultural landscapes targeted in this study (i.e. characterized by small-holder farmers, with low levels of management and crop densities). Altogether they all contribute to the fact that the relationship between crops on the ground and their spectral signature is less evident:

Geometric resolution: related to the pixels size. In this case parts of the cropped areas can be characterised by fields which are too small to be shaped by the pixels. This aspect could, at least in principle, be reduced by the introduction of EO data with higher geometric resolution.

Atmospheric conditions: different atmospheric conditions can occur, even at local level, when the dates of the images are distant in time.

Crop phenology: each crop speciesand even variety develops according to different phenological phasesto which representative spectral signatures can be associated.It is thus very difficult that crops reach the most representative spectral signature at same time.

Seeding date: even within a species and variety having a specific phenology, the starting (i.e. seeding) date of the crop cycle can change according to several factors including climate and, locally, topography and soil conditions. Management choices of farmers are also very important.

Continuous cropping: the agro-ecologic conditions allow some overlap in the growing cycles of crops especially in sub-tropical environments.

Intercropping: in the same parcel different crops are sown at the same time or with a little time lag (relay cropping) resulting in a non-distinctive spectral signature, being a mix between two existing classes. A similar problem is given in case of weeds.

Crop density: it relates to farmers management practises, soil fertility, pathologies, water stress. Soil in the background alters the natural spectral signature.

Difficulty in determining the best date to collect the satellite images: it is due to the unpredictability of the weather conditions and the lack of synchronization of cultural practices/crops phenology.

The area estimations can be now calculated based on the binary stratification described in section 2. The crops where the classification performs better are selected for this purpose, i.e. Groundnuts in Senegaland Maize in Kenya (Butere-Mumias only). Results are given in Table 4 and 5:

Table 4: *Estimations with and without the image classification: Groundnuts, Senegal*

Approach	Crop area (ha)	StdDeviation	CV
only survey	72.229.70	2.713.10	3.76
survey + classification	72.597.02	2.688.12	3.7

Table 5: *Estimations with and without the image classification: Maize, Kenya*

Approach	Crop area (ha)	StdDeviation	CV
only survey	29.537.73	2.633.56	8.92
survey + classification	29.421.20	2.615.51	8.89

Comparing the results above the gain in precision (expressed in terms of reduction of the CV) achieved introducing the classification in the area estimates can be observed. Since there is only a minor decrease in the CV for both areas and crops, one can conclude that the contribution of the image classifications to further improve the accuracy of the estimates from the surveyis altogether very modest.

A more formal way to assess the contribution of EO data can be expressed in terms of Relative Efficiency (RE), which is the ratio between the variance of the ground survey area estimate and the variance after this estimate has been corrected withthe aid of classified satellite images. RE values close to 1 indicate that the EO contribution towards reducing the sampling variance estimations is very low:

$$RE = (Var.withouteo \,/\, Var.with\ eo)$$

Where:

Var.without eo = Variance estimation obtained without the contribution of EO data
Var.with eo= Variance estimation with the contribution of the EO data.

In both areas the RE is in the order of 1.01 (1.009 for Senegal and 1.007 for Kenya).If the cost components are known, also the cost efficiency can be computed. This is expressed in terms of Net Relative Efficiency (NRE) and calculated as follows:

$$NRE = RE \ x \ (Costs \ without \ eo \ / \ Costs \ with \ eo)$$

Where:

Costs without eo = Costs without the contribution of EO data (i.e. ground survey costs).
Costs with eo = Costs considering the contribution of EO data (i.e. ground survey costs, cost of the images, of the radiometric and geometric corrections and of the classification).

In the two selected areas the NRE is in the order of 0.21, which indicates a rather low performance in terms of cost-efficiency.

4. Conclusions

As far as the generation of crop area estimates the contribution of EO was evaluated for two of the areas (and the main crops cultivated within) selected as use cases in the project AGRICAB. As to the contribution to the third level of application (image classifications to improve accuracy of estimates from the ground surveys) the results were evaluated in terms of RE and cost efficiency. The result indicates a low contribution in improving the estimates and, as a consequence and due to the type of images used, an even lower cost-efficiency.

Nevertheless, the other two levels of application, i.e. the construction and stratification of the sampling frame, and the support to ground surveys, were deemed very important, although it was not possible to further quantify their contribution.

In the future, in order to increase the efficiency in the use of EO data, a multi spectral/temporal approach can be further explored. A minimum of at least two images must be foreseen. A critical issue related to the areas of interest and similar environments is the presence of cloud cover during the main growing seasons. This poses a constraint on the proposed multi temporal approach using optical satellites only. In perspective, a combination of SAR (Sentinel 1) and optical satellites (Sentinel 2) may provide a useful opportunity for this type of applications.

REFERENCES

Carfagna, E. (2013) Evaluating the cost-efficiency of remote sensing in developing countries.FAO Global Strategy, first SAC meeting, Rome.

Ceccarelli, T. and Remotti, D. (2012)Service utility report: agricultural survey optimisation service in Malawi.ESA report. Paris.

Ceccarelli, T., Remotti, D., Haub, C. (2013)Agricultural survey optimization service: Malawi report on activities, 2013. ASO Malawi final meeting, Lilongwe.

FAO (2015)Technical Report onCost – Effectiveness of Remote Sensing for Agricultural Statistics in Developing and EmergingEconomies. Technical Report Series GO-09-2015, FAO, Rome.

Gallego J.F., Craig M., Michaelsen J., Bossyns B., Fritz S. (Editors), (2008) Best practices for crop area estimation with Remote Sensing. GEOSS Community of Practice Ag 0703a, JRC Scientific and Technical Reports.

Gallego, F.J., Carfagna, E. and Baruth, B.,(2010)Accuracy, objectivity and efficiency of remote sensing for agriculture statistics. in: *R. Benedetti, M. Bee, G. Espa& F. Piersimoni (eds.) Agricultural Survey Methods*, pp. 193–211. New York, John Wiley and Sons.

Giovacchini A., Remotti D., Monaci L., Downie M, Imala V., Mosquito Patricio F., Tauacale F., Diop B., (2015)AGRICAB AGRICULTURAL STATISTICS – TECHNICAL REPORTDeliverable 31.3.