DEVELOPING A SATELLITE BASED AUTOMATIC SYSTEM FOR CROP MONITORING: KENYA'S GREAT RIFT VALLEY, A CASE STUDY

Roberto Luciani[1], Giovanni Laneve[2], Munzer Jahjah[3], Collins Mito[4]

[1] DIAEE - Sapienza University of Rome, Via Salaria 00158, Roma, Italia, Email: roberto.luciani@uniroma1.it
[2] SIA - Sapienza University of Rome, Via Salaria 00158, Roma, Italia, Email: giovanni.laneve@uniroma1.it
[3] ASI - Agenzia Spaziale Italiana, Roma, Italia, Email: munzer.jahjah@asi.it
[4] Physics Department - University of Nairobi, Nairobi, Kenya, Email: collins@uonbi.ac.ke

ABSTRACT

The crop growth stage represents essential information for agricultural areas management. In this study we investigate the feasibility of a tool based on remotely sensed satellite (Landsat 8) imagery, capable of automatically classify crop fields and how much resolution enhancement based on pan-sharpening techniques and phenological information extraction, useful to create decision rules that allow to identify semantic class to assign to an object, can effectively support the classification process. Moreover we investigate the opportunity to extract vegetation health status information from remotely sensed assessment of the equivalent water thickness (EWT). Our case study is the Kenya's Great Rift valley, in this area a ground truth campaign was conducted during August 2015 in order to collect crop fields GPS measurements, leaf area index (LAI) and chlorophyll samples.

1. INTRODUCTION

Climate change is caused both by natural processes and anthropogenic activities; these activities include industrialization, deforestation or reforestation activities, agriculture and energy production, waste, urbanization and urban sprawl, land use change, soil erosion, drought and desertification, water degradation. Maintaining ecosystem functioning is a prerequisite for sustainable land management and planning of human activities: the territory knowledge plays a key role and the relevance of land monitoring and mapping, that finally leads to quantify changes in land cover, is widely recognized as essential in the study of global changes. Agriculture is the most important sector of many developing countries but long term productivity is threatened by increasingly intensive soil use for cultivation and livestock production. Agricultural land cover maps are critical for monitoring the current conditions and long term changes of crop and pasture lands. The UN (United Nations) FAO (Food and Agriculture Organization) Africover project generated the most recent and detailed land cover map for Central-Eastern African countries in the early 2000s. Since then, population growth and major policy changes have caused land use to shift to increased agro-pastoralism and systematic expansion of cropland area; on the other hand certain provinces of the country have been deeply affected by the lost of agricultural areas due to climatic change and human activities; moreover the burning of grass and bush savannah are resulting in loss of vegetation cover, that leads to the increasing of soil erosion and soil desiccation that can be worsen by drought and the declining of groundwater levels. All these considerations witnessed the need to updated land cover maps of cropped areas in order to better understand how much effective the land use shift could be in causing productivity loss and food security alarm for Kenya.

Automated, low-cost remote sensing tools are well suited for continuously monitoring crop growth and providing farmers with timely information about crop performance as vegetation indices derived from satellite imagery are well correlated with those parameters that defines the crop yield's status. The UN FAO Africover project was specifically dedicated to crop classification. After the Africover project, the ESA-GlobCover 2005 project delivered, to the international community in 2008, the first 300 m global land cover map based on MERIS FR (Full Resolution) images of year 2005. In 2010, the GlobCover chain was run by ESA and the Université catholique de Louvain (UCL) in order to derive a new global land cover map based on a 2009 MERIS time series. However, those land cover maps were characterized by a spatial resolution (300 m) which is not really suitable for monitoring agricultural areas especially in regions characterized by wide distribution of small farm holders such Kenya. The vegetation indices (VI) derived from satellite images are notoriously correlated with the parameters that define

Proc. 'Living Planet Symposium 2016', Prague, Czech Republic, 9–13 May 2016 (ESA SP-740, August 2016)

crops status and for nearly four decades the Normalized Difference Vegetation Index (NDVI) has remained one of the most consistently and widely measured vegetation indices across a wide variety of sampling platforms that provide us with plurality of information and historical archives; NDVI data at high temporal frequency have been widely used to track seasonal phenology of green-up and senescence over a wide variety of ecosystems from space using NOAA's Advanced Very-High Resolution Radiometer (AVHRR) [1,2,3 and 4] and NASA's Moderate Resolution Imaging Spectrometer (MODIS) [5,6]. As a result, the EO data are recognized as very important for monitoring the crops health status and productivity, and providing the necessary information to food security and early warning systems.

Our case study is the Kenya's Grain Basket (Rift valley); Kenya is the largest producer of tea in Africa and in 1999 it was the fourth in the world, behind India, China and Sri Lanka. Black tea is the main agricultural source of foreign currency in Kenya. Production in 1999 reached 220,000 tons. The tea exports were valued approximately 404.1 million USD in 2001, or about 18% of total exports: we choose the Kericho district's Tea plantations as a case study. Using Landsat 8 imagery, we set up an automatic monitoring system able to classify agricultural areas and detect land use changes [7]. Due to the high number of small farm holders the countryside is widely characterized by small and medium size fields (from 5 to 2 ha or even less) therefore the OLI (Operational Land Imager) Landsat 8 30 m spatial resolution could represent a significant limitation for the classification algorithm to carry out a proper image segmentation; to overcome this issue we have implemented the FIHS (Fast Intensity Hue Saturation) pan-sharpening technique [8,9] to increase Landsat 8 imagery resolution from 30 m to 15 m. A multi-temporal object-based classification technique was used for identifying the current agricultural area of interest. To improve the classification algorithm ability to clearly isolate the vegetation species currently under study a technique that implements both spectral and phenological information, previously extracted from OLI based NDVI time series, was used. Vegetation health status is deeply correlated with vegetation water content as it constitutes 40-80% of plant leaves volume [10]. Different wavelengths exhibit strong correlation with water content but this correlation can mainly depends on the magnitude and range of leaf sample water content and leaf geometry and structure: then different wavelength can be more or less appropriate in leaf water content estimate depending from vegetation species [11]. Different indexes for vegetation water content estimation are available in literature, one of the more often addressed is the Equivalent Water Thickness (EWT). EWT is the weight of water per unit of area and is calculated by using the Bowyer and Danson equation (2004):

$$EWT = \frac{FW - DW}{A}\,(\mathrm{g \cdot cm^{-2}}) \qquad (1)$$

The EWT is directly related to the depth of absorption in the SWIR bands and has a good correlation with leaf spectral parameters especially with water absorption feature in the range between 1400-2500 nm [12], and water absorption bands have been widely used for leaf water content assessment in several studies. It's worth to note that the depth of water absorption in the reflectance spectra seems not to be affected by leaf structures [13]. We are now investigating the possibility of remotely sensing EWT parameter over tea fields plantation. The overall goal of this study is to investigate the utility of introducing remotely sensed NDVI derived phenological information into a classification process in order to classify agricultural areas and to study the feasibility of retrieving EWT leaf information from a remotely sensed satellite based platform.

2. DATA & METHODOLOGY

2.1 Study area

The region of interest (ROI) is the Kericho county (Fig.1) located within the western highlands of the Kenyan great Rift Valley (White Highlands). This area is the home of the Kenya's bigger water catchment area, the Mau Forest; with a high altitude and virtually daily rains, Kericho county is the centre of Kenya's large tea industry. Tea cultivations are mainly rainfed, because of the proximity to Victoria Lake, the plantations benefit from the clouds natural formation caused by lake water evaporation. OLI Landsat 8 imagery are taken over a period that runs from April 2013 to May 2015 and they are characterized by a cloud cover contamination lower than the 30% of the entire image. Prosail 5B software has been used to perform reflectance simulation taking into account environmental settings taken from the ECMFW

(European Centre for Medium-Range Weather Forecasts) and from ground truth samples collected during the ground campaign.

Figure 1. Kericho County, Great Rift Valley, Kenya.

2.2 Pre-processing and resolution enhancement

An automated IDL based pre-processing tool devoted to the extraction of the metadata and georeference information for each acquired OLI image and to convert image Digital Numbers (DN) into radiance and Top of Atmosphere (TOA) reflectance values, has been implemented. The pre-processing unit is also dedicated to the final product spatial and temporal coordination: the acquired scenes are mosaicked to create maps at a local and regional scale; the scenes are also stacked together to organize a multi-temporal package. These two technique mixed together allow us to create OLI multi-temporal image mosaic at a 30 m spatial resolution. At this point we face the first issue: Kenyan countryside is mainly characterized by an high number of fragmented small and medium size farm holders that dramatically increase the classification difficulty; 30 m spatial resolution images are not enough detailed for a proper classification of such areas. A pan-sharpening technique has been implemented to increase image resolution from 30 m to 15 m; such techniques are well documented in literature for the ETM+ Landsat 7 imagery, but the resolution enhancement of OLI Landsat 8 imagery intended for vegetation studies is quite a challenge due to the lack of the near infrared band (NIR) in the OLI panchromatic channel. To overcome this issue the FIHS pan-sharpening technique has been applied and validated. The spatial enhancement is obtained by using the equation developed by Johnson in 2014:

$$MS_{high} = MS_{low} + (PAN - I)$$
$$I = 0.08B + 0.54G + 0.38R \qquad (2)$$

Where MS_{high} is the pan-sharpened multi-spectral band, MS_{low} is the original multispectral band re-sampled at the panchromatic image resolution by using the Nearest Neighborhood method (in order to reduce the arbitrary assumptions and preserve as much as possible the original image information content) and I is an intensity image derived from the multi-spectral bands (Blue, Green and Red channels) [7,8]; this new image is intended to be a simulated high resolution panchromatic image (based on the low resolution multi-spectral image): spatial resolution details are derived by subtracting the simulated panchromatic image from the original one then those details are directly added to the NIR low resolution band. The NIR band histogram, before and after the resolution enhancement, shown in Fig. 2, is taken into account in order to validate the FIHS technique:

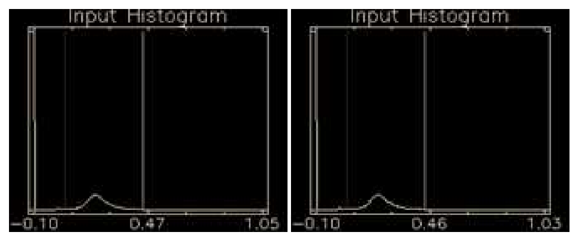

Figure 2. NIR histograms before (above) and after (below) the FIHS resolution enhancement

as clearly shown in the figure, histogram's shape is preserved during the transformation, this means that the DN values distribution is well preserved. It is worthwhile to note that the FIHS is an additive technique. If we consider to introduce the new high resolution NIR bands into a computational process, as the NDVI computation, the terms from which the resolution enhancement depends are well preserved during the process and not mutually erased as it will be the case if multiplicative pan-sharpening technique are used. Moreover, the resolution enhancement strongly depends by the Red-NIR bands difference and it will be as much higher as this difference increases; the algorithm is particularly suitable to be applied on vegetated areas where the gap between Red and NIR bands is higher.

2.3 Phenological information extraction

Phenology is the study of periodic plant and animal cycle and how these are influenced by seasonal and interannual variations in climate as well as habitat factors such elevation. The NDVI time series is staked into a multi-layer pseudo image in which each single band relates to an NDVI specific date (Fig.3). Ground test sites have been previously selected, searching for agricultural vegetated areas and to pave the way to information extraction. Phenological profiles provide historical NDVI projection over a specific site. They are little affected by cloud contamination and very useful in discriminating between different crop species, crop rotation cycles and biomass seasonal trends. With few exceptions represented by evergreen cultivation like tea or biennial cultivation like sugarcane, for each test site it is possible to build the crop sequence that has taking place and to identify the crop rotation and the preservation technique.

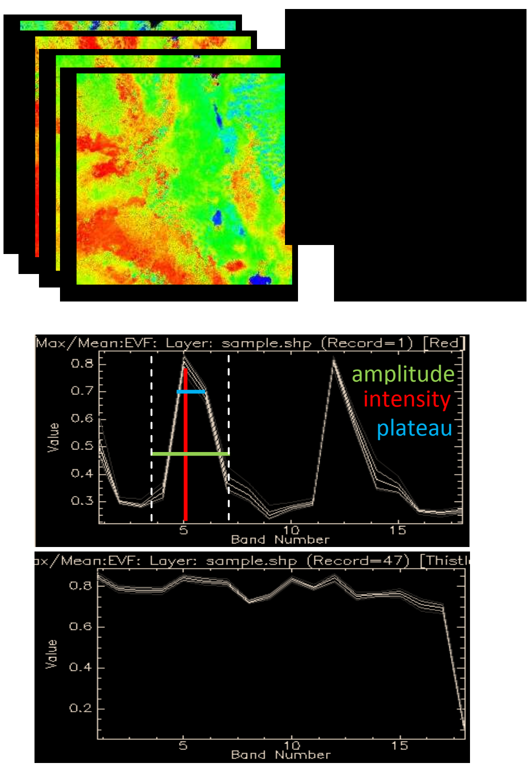

Figure 3. Phenological profile construction (above) and several samples extracted from the Kericho county during the referencing period (below).

The identification of the land use category defines the semantic classes and for each vegetated species the cultivation development characteristic, a temporal parameter that is related with phenological amplitude (Fig. 3). The maturation and health status peak are related with phenological intensity (Fig. 3). The presence of a phenological plateau could be indicative of the period for which the full maturation is preserved (Fig. 3). All these parameters allow the algorithms to recognize and group together similar crop species and to compare the annual developmental cycle with the historical one, which could allow to assess the crop reaction to stress conditions. Up to now 14 phenological classes out of the sampled phenological profiles have been recognized, included the one for tea fields.

2.4 Classification

A detail of a traditional unsupervised ISODATA classification of the Kericho district, in Fig.4, clearly shows its limits in discriminating different agricultural areas. The high variability of the ground cover and the fragmented nature of the countryside lead to several misclassification errors; forests areas and evergreen cultivation, both characterized by high values of the NDVI index, are often confused. It should be noted that agricultural areas during crop senescence and soil preparation period show no vegetal cover; a proper classification technique has to take into account this aspect that is complicated by variability introduced by

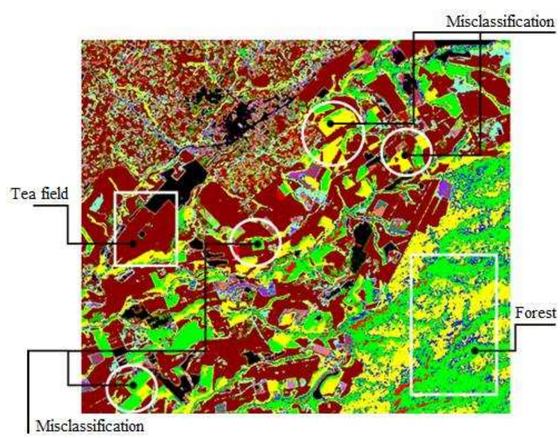

Figure 4. ISODATA unsupervised classification near Kericho Town; several misclassification errors between forested areas and tea fields are highlighted.

crop rotations technique and eventually the coexistence on the same area of different crop species at different developmental stage.

In order to reduce misclassification errors, several geographical maps were introduced to reduce the area where the classification process should be carried out. Africover, GLOBCOVER and Global Forest / Non-Forest ground cover maps have been manually updated using a supervised macro-categories classification map, NDVI and NIR false color image to build urban areas, water basins and forest areas masks. The resolution enhancement has made possible the identification of sample areas on the ground characterized by well developed transformation or preservation during the last decade. The sample areas from the dataset can be used like training or validation sites for the classification algorithm. For each single area, ground cover data and information about the land use have been collected and also validated through a field campaign carried out during August 2015. A further step was the introduction of the Bare Area Index (BAI):

$$BAI = \frac{band7 - band6}{band7 + band6} \tag{3}$$

Once the soil reflectance has been determined we can use it to exclude those no vegetated agricultural area from the single date classification. Nonetheless a single date classification has good probabilities not to show spectral differences enough significant to allow the algorithm to discriminate for different crop types. The NDVI index could be used as a discriminating factor to create separable vegetation classes due different health status and growth stage and then running separate classification process. The NDVI based segmentation means to introduce biomass information into the classification process; by the way this approach transfers much more information to the classification algorithm as we need to introduce all the available NDVI images for a specific site. Then we introduce phenological profiles to help the algorithm in discriminating for different crops and reducing the ambiguity related with the semantic interpretation of ground covers that can, possibly, show similar spectral signatures. This represents a kind of multi-temporal approach into a single date classification with the aim to reduce the problem semantic dimension; this reduction is due to the fact that, referring to a specific geographical zone, agricultural areas cannot show NDVI peaks (that is, a full maturation stadium), at any time of the year (with exception for evergreen cultivations), but only during characteristic periods. It should be noted that this approach is effective only for limited geographical regions, and this is especially true for all the regions characterized by sub-equatorial climate, that is responsible for the high variability in planning of the agricultural activities. For such areas ground truth data and interviews collection are really useful to asses remote sensed phenology information. For Central Africa regions the classification is forced into districts and provinces boundaries. Moreover we can impose now decisional rules outside the classification analysis and based on previously knowledge of the sample sites phenological characteristic, despite what we can do in a more traditional classification problem in which exclusion rules are exclusively internally generated with no relation with the physics of the problem and no control over them. Finally, a new classification process was carried out for the Kericho county with the aim to detect tea plantation: the result is shown in Fig. 5. The reduction of misclassification errors that, for the case study, mainly occurs between classes characterized by higher NDVI values, leads to a significant improvement of the classification performance; red polygons, which represents tea fields, are superimposed over the old Africover classification: Africover's polygons doesn't represents fields boundaries but only thematic areas to which a codified legend associates ground cover percentage. The updated classification map is capable of showing crop fields boundaries, cultivations total extent and exact location. All the classified areas fall into Africover polygons for which a tea fields ground cover percentage ranging from 40 % to 80 % has been indicated. Machine computational burden is reduced, as well as the time needed to carry out the segmentation. Once the phenology is known it is possible to select the salient features that significantly describes the associated semantic classes. A phenology based classification process needs the a priori definition of the semantic categories. This means that a previous study of the area of interest is necessary to make the classification properly run. However, if a semantic class has not been loaded in the machine, the associated crop type would not be recognized. In this case the agricultural area remains unclassified and it is necessary to revise the crop distribution on the ground. We can take advantage of such a characteristic thinking about a monitoring system: if the initial agricultural ground cover is well mapped, every new unclassified area leads to a change detection.

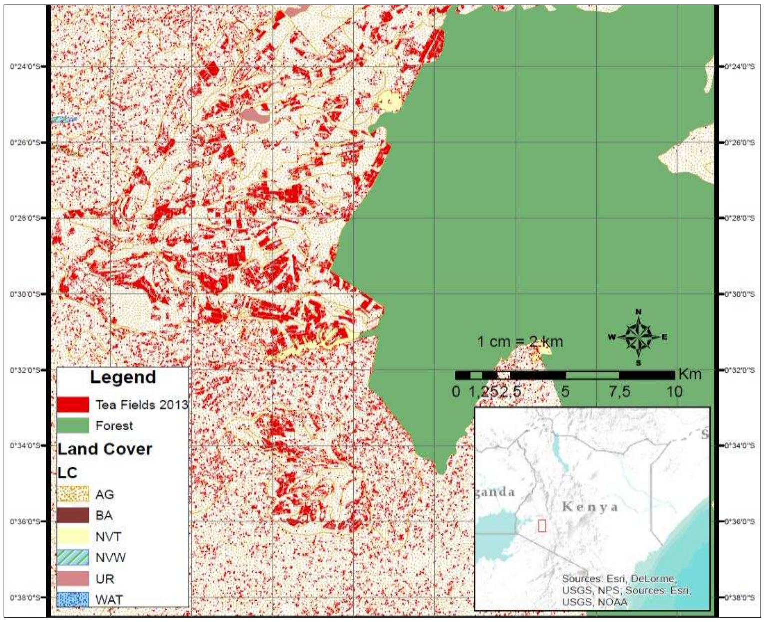

Figure 5. Detail from Kericho county Tea fields classification for the year 2013 (Tea plantation in red).

This aspect need to be furthermore investigated aside the main development that, for now, will consist in adding new features in order to be capable to discriminate the highest possible number of crop types within a single agricultural area of interest.

2.5 EWT evaluation

Vegetation health status is deeply correlated with vegetation water content and this is a critical parameter in evapotranspiration evaluation. The possibility to estimate the EWT by means of remote sensed platform can represent an important improvement for every system requesting the estimation of evapotranspiration data, due to the fact that EWT determination is not expensive even if time consuming. Several simulations based on Prosail 5B have been performed to understand the EWT dependence from typical vegetation parameters and then assess the feasibility of a tool capable of estimating the vegetation equivalent water thickness starting from the vegetation reflectance signature. Simulations involve chlorophyll content, carotenoid content, brown pigment content, LAI, leaf structure and geometry of observation variations. They have demonstrated how LAI variations can interfere with EWT evaluation, while all the others parameters are not affected by significant change due to EWT variation or are not synchronously varying with it. Then we carried out a specific set of simulations, taking into account only OLI Landsat 8 channels, during which EWT-LAI couple of values were varied simultaneously. We detect two trends: certain channels (like the one at 635 nm) cannot discriminate EWT-LAI couple of values since it shows a wide indetermination for each reflectance value. As shown in Fig. 6. in this case it’s not possible to discriminate for different EWT-LAI pairs. On the other hand channels, like the one at 1373 nm, isolate different EWT values for the same LAI; an ambiguity still exists between different and consecutive LAI values; this issue can be overcome using the channel at 1600 nm or LAI measurements previously collected. LAI samples collected throughout the Kenya’s Rift Valley were used to validate the preliminary study, finding out that the OLI Landsat 8 at 635 nm, 1373 nm and 1600 nm are very effective in detecting the EWT sample value.

3. CONCLUSIONS

The Italian Space Agency (ASI) and the Università di Roma 'La Sapienza' have recently signed an agreement concerning the activities to be carried out at Broglio Space Center (BSC) of Malindi (Kenya). In the framework of this agreement ASI is going to fund some research/training projects on topics of interest for both Italian and Kenyan institutions.

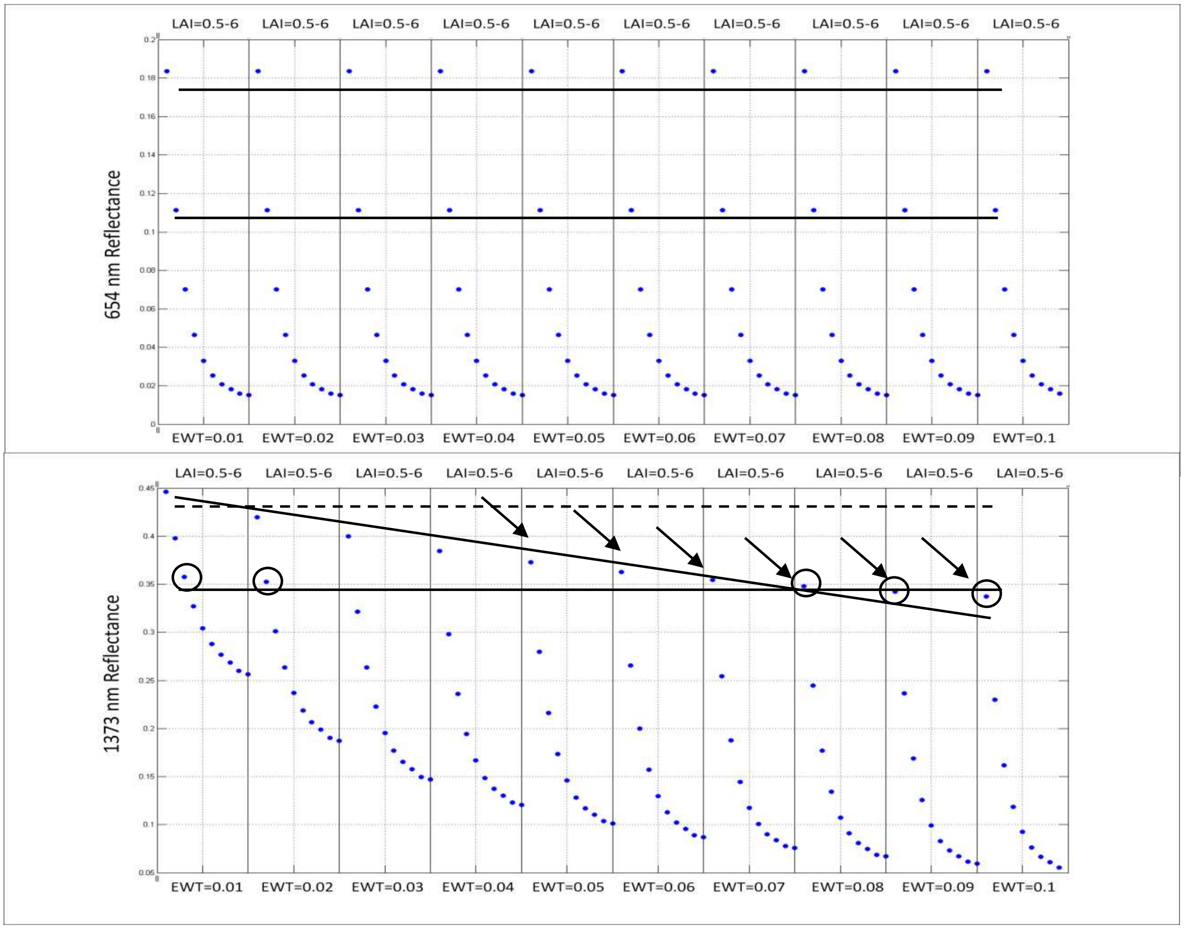

Figure 6. Prosail 5B simulations: in the first case (654 nm channel, above) is not possible to discriminate for different EWT-LAI couples; in the second case (1373 nm channel, below) the same LAI value is associated with different EWT value, but an ambiguity still exists.

This paper is devoted to describe the first year of activity carried out in the framework of one of the projects funded by ASI. The three years project 'System Implementation and Capacity Building for Satellite- Based Agricultural Monitoring and Crop Statistics in Kenya (SBAM)' aims at: developing a validated satellite based methods for estimating and updating the agricultural areas in the region of Central-Africa; developing methods and products allowing the assessment of the crops status in test areas (Kenya) by combining ground and satellite data; implementing an automated process chain capable to periodically provide agricultural land cover maps of the area of interest and possible an estimate of the crop yield. During the first year of the project a classification scheme, based on spectral and time series of satellite images, has been developed capable to identify different crop types. This represents a significant improvement of the most accurate land cover/land use map presently available for the country of interest (Kenya), that is the maps produced in the framework of the Africover project in the year 2000. The preliminary study about the feasibility of remotely estimating vegetation health status encourage us to pursue this line of research.

4. REFERENCES

1. Justice, C.O., Townshead, J.R.G., Holben, B.N., Tucker, C.J.(1985). Analysis of the phenology of global vegetation using meteorological satellite data. *International Journal of Remote Sensing*, 6, 1985:1271-1318

2. Justice, C.O., Holben, B.N., Gwyne, M.D. (1986). Monitoring East African vegetation using AVHRR data. *International Journal of Remote*

Sensing, 7: 1453-1474

3. Justice, C.O., Hiernaux, P. (1983). Monitoring the grassland of the Sahel using NOAA AVHRR data: Niger 1983", *International Journal of Remote Sensing*,7,1475-1498

4. Vrieling, A., de Beurs, K.M., Brown, M.E. Variability of African farming systems from phenological analysis of NDVI time series, Climatic Change DOI10.1007/s10584-011-0049-1

5. Fisher, J.I., Mustard, J.F. (2007). Cross-scalar satellite phenology from ground, Landsat, and MODIS data. *Remote Sensing of Environment* 109 261–273

6. Hmimina G., Dufrêne, E., Pontailler, J.Y., Delpierre, N., Aubinet, M., Caquet, B., de Grandcourt, A., Burban, B., Flechard, C., Granier, A., Gross, P., Heinesch, B., Longdoz, B., Moureaux, C., Ourcival, J.M., Rambal, S., Saint André, L., Soudani, K.. (2013). Evaluation of the potential of MODIS satellite data to predict vegetation phenology in different biomes: An investigation using ground-based NDVI measurements. *Remote Sensing of Environment* 132 145–158

7. Johnson, B. A., Scheyvens, H., Shivakoti, B.R. (2014). An ensemble pansharpening approach for finer-scale mapping of sugarcane with Landsat 8 imagery. *International Journal of Applied Earth Observation and Geoinformation* 33 218–225 8.

8. Johnson, B. (2014). Effects of Pansharpening on Vegetation Indices. *ISPRS Int. J. Geo-Inf.* 3, 507-522

9. Luciani, R., Laneve, G., Jahjah, M., Mito, C. (2016). Developing a classification method for periodically updating agricultural maps in Kenya. *Proceeding of IEEE Young Professional Conference on remote Sensing*.

10. Shen Y, Shi R, Niu Z, Yan C (2005) Estimation models for vegetation water content at both leaf and canopy levels. *Geoscience and Remote Sensing Symposium, 2005. IGARSS '05*. Proceedings. 2005 IEEE International 2:1387-1389

11. Bowyer P, Danson FM (2004) Sensitivity of spectral reflectance to variation in live fuel moisture content at leaf and canopy level. *Remote Sensing of Environment* 92:297-308.

12. Zhang J, Xu Y, Yao F, Wang P, Guo W, Li L, Yang L (2010) Advances in estimation methods of vegetation water content based on optical remote sensing techniques. *Science China Technological Sciences* 53:1159-1167.

13. Danson FM, Steven MD, Malthus TJ, Clark JA (1992) High-spectral resolution data for determining leaf water content. *International Journal of Remote Sensing* 13:461-470.