ASSESSING FIELD-LEVEL MAIZE YIELD VARIABILITY IN TANZANIA USING MULTI-TEMPORAL VERY FINE RESOLUTION IMAGERY

STEPHEN KIBET
FEBRUARY, 2016

SUPERVISORS:
Dr. Ir. A, Vrieling
Dr. R. Zurita Milla

ASSESSING FIELD-LEVEL MAIZE YIELD VARIABILITY IN TANZANIA USING MULTI-TEMPORAL VERY FINE RESOLUTION IMAGERY

STEPHEN KIBET
Enschede, The Netherlands, FEBRUARY, 2016

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfillment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialization: Natural Resource Management

SUPERVISORS:
Dr. Ir. A Vrieling
Dr. R. Zurita-Milla

THESIS ASSESSMENT BOARD:
Prof. Dr. A.D. Nelson (Chair) NRS Department, ITC
Dr. J. (Jan) Dempewolf (External Examiner) University of Maryland, USA)

ABSTRACT

Timely and accurate information on crop production is important for planning food-related decisions at both government and household level. However, acquiring such data is often a major challenge in most countries in Sub-Saharan Africa. The crop fields in these countries are highly fragmented with fuzzy boundaries and a complex cropping system. The use of coarse spatial resolution imagery (> 250 meters) in such landscape is often limited by mixed pixel problem and mismatch between field boundaries and the image pixel size. However, rapid technological development has seen improvement of remote sensing technologies whereby acquisition of very fine spatial resolution imagery (< 1 meter) with improved revisit time of less than a day, has been made possible at affordable cost. Such imagery include Unmanned Aerial Vehicle (UAV) and satellite data such as WorldView (WV) provided by Digital Globe (DG). These high quality remote sensing products have wide range of applications in many fields including agriculture.

This study was a proof-of concept to determine applicability of fine spatial resolution data in improving maize yield estimation at field level. The study was conducted in Kilosa District, Tanzania. The main aim of this study was to estimating maize yield at field-level using fine spatial resolution UAV, WorldView -2 and WorldView-3 images. Vegetation index metrics (VI) were derived from these fine spatial resolution images and together with field-level interview yield data, an empirical linear regression models were developed. Availability of same date UAV and WV images provided an opportunity to test performance of VI derived index by integrating the two datasets. Bootstrap statistical technique was applied in model validation. The optimal model with high adjusted coefficient of determination (adjR^2), low Root Mean Square Error (RMSE) and low standard error (SE) was used to derive yield variability map. The resulting yield variability map was correlated with field collected maize yield data using Spearman's rank correlation in order determine the relationship between spatial yield variability map and the actual yield status.

Results indicate that the Enhanced Vegetation Index (EVI) outperformed the popularly used Normalized Difference Vegetation index. EVI explained 63% of maize yield variability. The optimal period was found to be at fruit development stage of maize growth which occurs 60-75 days after sowing. The single-date VI showed to be the best predictor, followed by cumulative VI (cumVI) while maximum VI (maxVI) explained the least variability. In terms of the sensor performance, WorldView outperformed UAV as it had consistently large R^2 with maize yield. The correlation between same date UAV and WV showed a good correlation of R^2=0.51 using randomly selected averaged NDVI values. However, result of new WV NDVI derived from UAV using the linear equation computed from same data UAV and WV gave an R^2 of 0.44 indicating good potential of fusing VI data acquired from UAV and WV data. The yield variability within the fields had a coefficient of variation of 33%. In terms of the effect of field management factors on yield, weeding and method of tilling showed to have a significant impact on yield. Although high correlation coefficient was realized with the single-date imagery, most of the other metrics apart from cumulative vegetation index showed a weak relationship with yield. Furthermore, a scatter plot derived from the maize yield model showed an unusual trend where for high yield, it corresponded to low EVI. As a result of this, it was noted that the study did not give convincing results as to the performance of fine spatial resolution in estimating yield as it was limited by high differences in field management practices.

Keywords: Maize yield, UAV, WorldView, High resolution, spatial, variability and management factors

ACKNOWLEDGEMENTS

Glory be to God for his provision, care, and protection during my entire study period.

I am heartily indebted to ITC through Spurring a Transformation for Agriculture through Remote Sensing (STARS) project for funding my MSc studies here in the Netherlands. I wish also to pass my sincere gratitude to Dr. Rolf de By, (ITC) for noting the potential in me and for your continuous encouragement in my academic journey.

Words cannot express my gratitude to my supervisors, Dr. Ir. Anton Vrieling and Dr. Zurita Milla for your constructive criticism, ideas and always being readily available to advise during my entire MSc thesis period. Working with you, I have not only learned how to be a critical thinker, but also I have been inspired and learned a lot from your diligent work. I also wish to thank the chair during proposal stage, Dr. Ir. Kees De Bie for providing necessary backstopping in my work. My further appreciation goes to ITC staff who laid a firm foundation during the course work period which has enabled me undertake this research successfully.

To the STARS project partners in Tanzania which include Sokoine University team led by Prof. Siza Tumbo and Dr. Sixbert Maurice, I acknowledge the great logistic and data support granted to me during my fieldwork including. I also wish to extend my gratitude to the driver, Mr. Alex who not only made sure I was on schedule, but also shared with me words of wisdom. Further appreciation goes to the University of Maryland team lead by Dr. Jan Dempewolf, for setting a good landing ground for my fieldwork and for providing pre-processed UAV images. To the ITC staff, Dimitris, thanks a lot for your quick action in providing access to aerial and satellite images, not forgetting great support in pre-processing the WorldView images. I am also grateful to the field guides, led by Mr. Mogera, agricultural extension officer in Kilosa and the local elders for facilitating my field interviews with the farmers. Your perseverance during the long hours of working in the scourging sun was a great encouragement.

The NRM and GEM class of 2014/2015, you are a precious gemstone, thanks for providing social and academic support. The Kenya team, including the East Africa community students from Rwanda, Tanzania, and Uganda, our interactions made me feel at our motherland Africa, though miles away. Specifically I wish to acknowledge Mimy, Maurice, Aristotle, Zemeron, Lydia, Agnes and Edson for your support.

To my dad, brother, sister, Naomi, Nicole and lovely nieces and nephews, 18 months seemed a long time, but with your constant support, prayers and encouragement I have made it. Although the unfortunate happened when we lost our beloved mum during my study period, her zeal to educate us is still burning within me and this MSc. is a special dedication to her tireless effort to educate us. RIP mum, we love you and we are here to carry on the mantle and legacy you left behind.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1. Introduction

Agriculture plays a significant role in achieving the World Bank Group agenda of ending poverty and hunger by 2030 (Townsend, 2015). Globally, 805 million people are estimated to be chronically undernourished, of which 23.8 % live in sub-Saharan Africa (FAO et al., 2014). To improve this situation, the World Bank (2008) highlights the importance of agriculture and its related industries as a principal option for spurring growth, overcoming poverty and enhancing food security in the Sub-Saharan Africa (SSA) region. In this predominantly agriculture-based economy, small-scale farmers account for 75 % of the region's agricultural production and 75% of employment (Salami et al., 2010)

In East Africa, maize (*Zea mays*) is an important cereal food crop planted annually on approximately 7.3 million hectares corresponding to 21% of the arable area and 41% of the land under cereals (Erenstein et al., 2011). It is typically rain-fed and is cultivated across a range of latitudes, altitudes, moisture regimes, slopes and soil types (Livingston et al., 2011; Smale et al., 2003). Maize is primarily produced for home consumption and for local markets by small-scale family farms (Erenstein et al., 2011). In Kenya and Tanzania, maize consumption represents on average 40% of the daily dietary calorie requirement (Groote et al., 2002).

Maize yield in the region shows a high spatial and temporal variability. Large-scale spatial variability can be explained by differences in rainfall and soil characteristics (HarvestChoice, 2010; Marques da Silva et al., 2008; Smale et al., 2011; Thornton et al., 2009; Yengoh, 2012) while small-scale variability is importantly influenced by farm management decisions like sowing dates, weeding, pests, diseases, fertilizer application and method of tilling applied. Furthermore, small-scale variability is attributed to biophysical factors such as rainfall, soil properties, elevation and floods (Nathan, 2014; Sacks et al., 2010; Vyas et al., 2013). An important determinant of temporal variability of maize yields is the interannual variability of rainfall and temperature, resulting in frequent droughts in the region (Funk et al., 2009; Magehema et al., 2014; Porter et al., 2005). This large yield variability underlines the need to assess and monitor yields within the growing season.

Maize yield can be obtained by dividing maize production by the cultivated (or harvested) area. Data on maize production and area cultivated are often derived from area frame sampling and statistical farm register (Everaers, 2010). Area frame sampling is the breakdown of a land area into relatively homogenous sampling units commonly referred to as primary sampling units (PSU) (Willett, 1981). Aerial photographs and remote sensing images such as Landsat has been used in dividing these areas upon which farmers interviews are carried out. Although area frame sampling is a well-developed and efficient technique for collecting agricultural data, it is limited by high cost and tedious implementation process. The second approach is the use statistical farm registers. These refer to up-to-date agricultural registers kept by the government ministries at a different administrative level which includes household demographics, market information, business and tax registers. Upon compiling all these sources of data, detailed agricultural statistics at the household level can be obtained at relatively low cost. However, one major challenge with farm registers is linking registers with different variables can be tedious and also there is the issue of accuracy of information provided in these records (Turtoi et al., 2012; Väisänen, 2009).

Remote Sensing data has a wide range of applications in the field of agriculture. Some of these applications include maize yield estimation (Claverie et al., 2012; Lewis et al., 1998; Nathan, 2014; Prasad et al., 2006), crop mapping (Jain et al., 2013; Khan et al., 2010), source input data for crop models (Reynolds et al., 2000; Vintrou et al., 2014) and as an indirect indicator of crop yields (Van Wart et al., 2013). Indirect indicators are usually obtained by evaluating the inter-seasonal variability of vegetation indices derived from coarse-resolution (>100m) optical sensors, and empirically relating these to measured crop yield (Funk et al., 2009; Rembold et al., 2013; Wu et al., 2013). This is based on the premise that crop yield strongly relates to the green biomass which develops over the season and which is often estimated from spectral properties derived from satellite observations (Meroni et al., 2013).Examples of such indices include Normalized Difference Vegetation Index (NDVI) which is the widely used vegetation index, but many other indices exist that have been used to better account for atmospheric and soil background effects (Henrich et al., 2012; Qi et al., 1994). The availability of dense time series remote sensing data from coarse resolution images has been exploited to derive time-related vegetation index metrics (VI) data and applied to crop yield estimation (Bolton et al., 2013; Wang et al., 2014). These time-related VI metrics commonly referred to as phenology metrics describes the timing of vegetation events using data derived from synoptic sensors (Brown et al., 2008, 2010; de Beurs et al., 2005).

Although there are a number of vegetation metrics that has been applied in vegetation studies, this study will focus on three specific VI metrics, single-date vegetation index, cumulative variable vegetation index (cumVI) and season's maximum vegetation index (MaxVI). A key rationale for using coarse-resolution data in most yield assessment studies is their short (daily) revisit time with global coverage, which permits to precisely follow vegetation development even in the case of frequent overcast conditions and to reduce atmospheric effects (Atzberger, 2013; Rembold et al., 2013).

Although coarse-resolution time series data provide relevant input for assessing crop production, a number of limitations exist. Coarse spatial resolution measurements of spectral reflectance contain mixed information from several surface types hence complicating signal interpretation. Moreover, with coarse resolution data, it is difficult to classify specific crop types given most crop fields in SSA are small and regularly multi-cropped (Lobell, 2013; Nathan, 2014; Rembold et al., 2013). Besides the small agricultural parcels giving rise to mixed spectra, crop condition and yields can also vary widely between fields making it difficult to directly relate a spectral or temporal signature to a specific crop occurrence or condition (Hoefsloot et al., 2012). In order to avoid mixed pixels problem, Claverie et al. (2012) suggested the use of fine spatial resolution data (<10 meters). However, finer spatial resolution mostly implies a lower observation frequency and a high cost.

The recent development of sensors collecting fine spatial resolution data at shorter temporal intervals is opening a new frontier in agricultural monitoring. The mixed pixel limitation from coarse spatial resolution remote sensing data is being progressively reduced by the availability of fine spatial and temporal resolution sensors (Johnson et al., 2012; Rembold et al., 2013). While one avenue could be to combine information from fine and coarse resolution sensors using image fusion techniques (Gevaert et al., 2014; Stenger et al., 2009; Laigang Wang et al., 2014; Zurita-Milla et al., 2011), new fine-resolution satellites are being launched that directly provide shorter revisit capabilities. For example, the Sentinel-2A satellite launched on 23 June 2015 is capable of monitoring variability in land surface conditions due to its wide swath width, 13 multi-spectral bands in visible, near infrared and shortwave spectrum coupled with a high revisit time of 5 days once Sentinel-2B is in place 2016 (European Space Agency, 2015).

In parallel to satellite developments, the use of airborne sensors such as those onboard Unmanned Aerial Vehicles (UAV) is increasingly being adopted for crop monitoring and yield assessment (Geipel et al., 2014; Lin et al., 2011; Palermo, 2015). UAV sensors provide very fine resolution data of up to 1 cm depending on the flight height, camera type, and sensor resolution with flexible revisit time as determined by the user (SenseFly Ltd., 2015). A good example is a study by Geipel et al., (2014) where they combined crop height model with fine resolution VI derived from UAV RGB bands and which was able to explain 74% of maize yield variability. Fine spatial resolution imagery is important for establishing better maize VI-yield relationship at early stages of crop development, which gets less important as the crop grows to a point of becoming disadvantageous (Geipel et al., 2014). This is attributed to high soil reflectance during early growth stages which reduce progressively as the crop grows. The increasing use of fine spatial and temporal sensors is driven by the need for accurate field level monitoring and the growing need for micro-level planning (de By et al., 2015; Singh et al., 2002).

Despite the promise of satellite and UAV data of fine spatial and temporal resolution for crop yield estimation, until the present, only a few studies have been carried out that reliably estimate yields. Particularly for smallholders systems in East Africa, it is envisaged that important advances could be made in accurately estimating maize yield at field-scale from very fine spatial resolution and multi-spectral imagery. An initial approach to achieve this is to evaluate if VI-maize yield empirical relationships can accurately describe the link between VI and field-level yield data at the different moment of the season. If feasible, such relationships could potentially be extrapolated to obtain yield estimates for larger areas. The study, therefore, aimed at establishing an optimal vegetation index (VI) and best period for estimating field-level maize yield using fine resolution UAV and WorldView imagery and develop maize yield spatial variability map which would be explained based on filed-level management information collected during the field work. This study is carried out in the context of Spurring Transformation in Agriculture through use of Remote Sensing (STARS), a project which is led by Faculty of ITC, University of Twente, in partnership with five other leading research organizations, private companies, local research institutes and government ministries in East Africa, West Africa and South East Asia.

The main research objective of this MSc thesis is to study maize yield variability from fine spatial resolution, multi-temporal Unmanned Aerial Vehicle (UAV) and WorldView (WV) imagery and explain this variability from differences in field management for two 1 x 1 km areas in Kilosa district, Tanzania.

To achieve this, the following specific objectives are defined:

1. To establish empirical relationships between field-level interview maize yield data and UAV/WV derived vegetation indices (derived from single-date and multi-temporal images)
2. To apply the empirical relationship that explains most of the yield variability to the two 1x1 km areas to visualize spatial differences in maize yield;
3. To determine in-field spatial yield variability using field interview data and explain the variability based on differences in field management practices;

In order to achieve these objectives, the study was guided by the following research questions:-

a) Which vegetation index metrics and timing explain most of the yield variability as derived from single-date, cumulative VI and maximum VI metrics?
b) What is the maize growing stage for the optimal maize yield assessment using VI metrics and field interview data?
c) To what extent does estimated maize yield vary within and between fields?
d) Can we discern maize fields that clearly show a high, average low maize yield variability
e) To determine in-field spatial yield variability using field interview data and explain effect of management factors on yield.

2. STUDY AREA AND DATA

2.1. Study Area

The study was carried out in two 1 km by 1 km sites in Gongoni and Mbuyuni locations in Kilosa District, Morogoro Region, Tanzania (37.122 E; 6.652 S and 37.142 E and 6.672 S) as shown in Figure 1. The elevation ranges between 350 and 500 meters above sea level with a sloping rising of less than 10 percent.

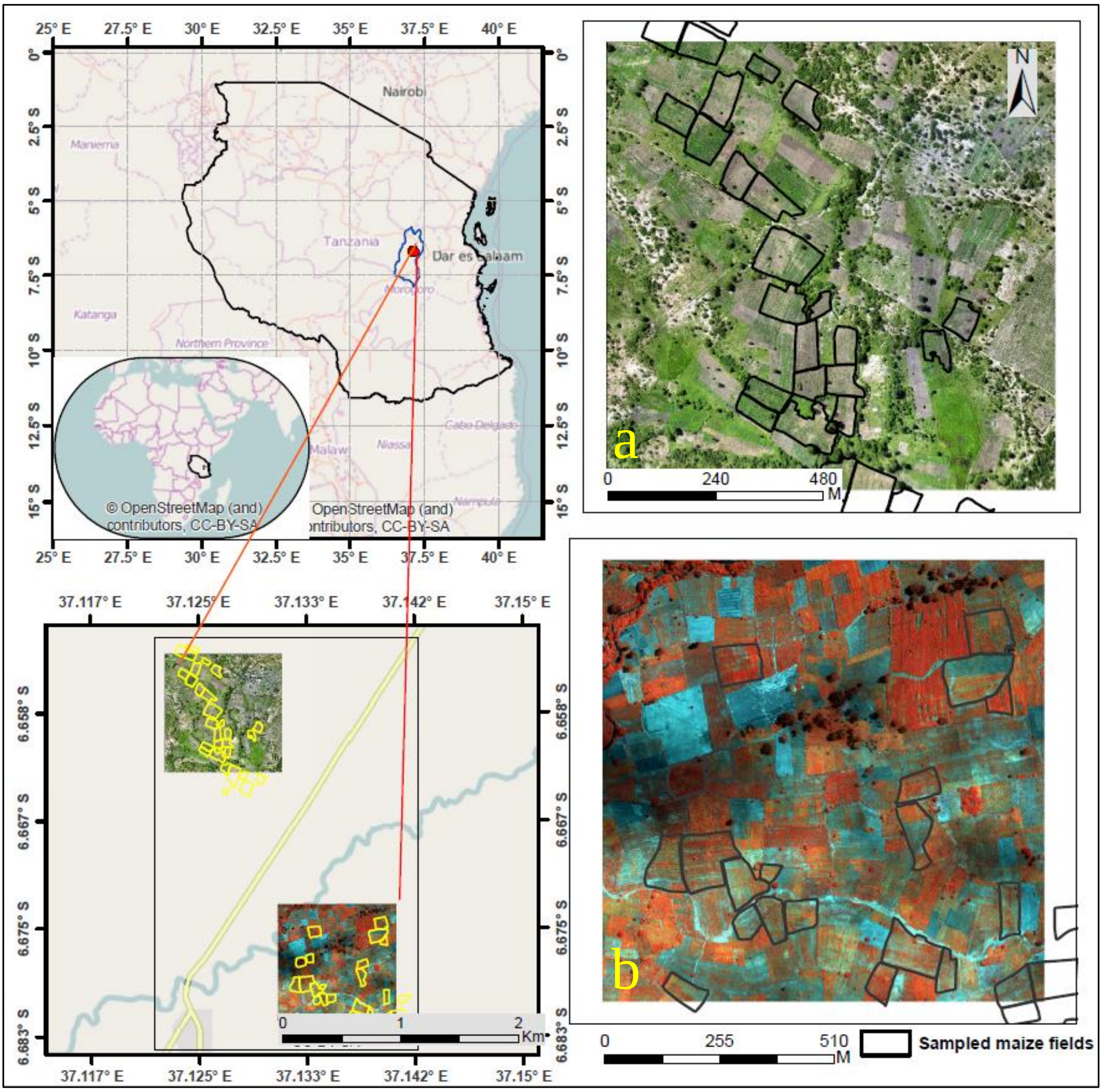


Figure 1: The location of the 1 km by 1 km field study sites: (a) Gongoni site imaged by a true-colour UAV of 19 April 2015; and (b) Mbuyuni site imaged by a false-colour UAV image of 19 April 2015. The black lines indicate some of the sampled maize fields.

Kilosa District has an average annual rainfall of 976mm per year divided by two rainfall seasons. The main rainfall season starts in February to June with May being the wettest month. The district experiences an average eight months of rainfall (October-May), with the highest levels between February and March. The rainfall distribution is bimodal in good years, with short rains (October-January) followed by long rains (mid-February-May). However, the year 2015 and the previous two years seem to show a different trend with rainfall pattern indicating a single rainfall season based on agro-climatic condition monitor developed by Group on Earth Observations Global Agricultural Monitoring (GEOGLAM). The rainfall data is based on 0.05 degree resolution 10-daily rainfall regional average estimates from Climate Hazards Group Infrared Precipitation with Station data (CHIRPS) and NDVI composites from MODIS data (Moderate Resolution Imaging Spectro-radiometer); which is an e-MODIS product of the United States Geological Survey (USGS) acquired by the Terra satellite and consist of 8-day maximum value NDVI composites at 250 m resolution. The mean annual temperature is 24.6 °C with a daily mean maximum temperature of 26.9°C during the rainfall season in the month of May and lowest of 21.8°C during the dry months of July and August.

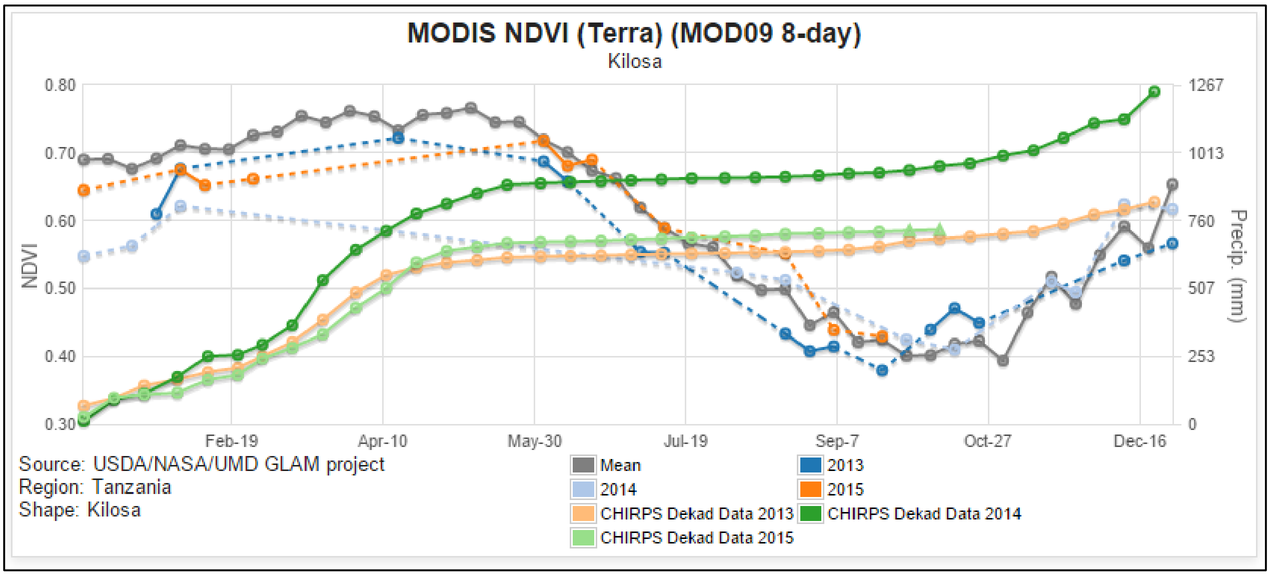


Figure 2: Kilosa District 10-day rainfall estimates and 8-day NDVI time series composites over the period 2013-2015 and mean NDVI for the year 2000-2015.

The soil in the area has varying proportions of sand, silt and clay as presented on the soil maps in Figure 3. The map was derived from International Soil Reference and Information Centre (ISRIC) Soil Information database. The soil data comprises 250 m global soil database with different characteristics modeled from satellite-derived data and validated with more than 3000 ground sample points (ISRIC - World Soil Information, 2015). The dominant soil texture in Gongoni is sandy clay while clay is dominant in Mbuyuni (described using online soil texture pyramid developed by United States Department of Agriculture (USDA). The difference in soil texture was evident within and between fields with varying color differences; sandy soils having dominant bright colors while clay soil having dark in color (Figure 4).

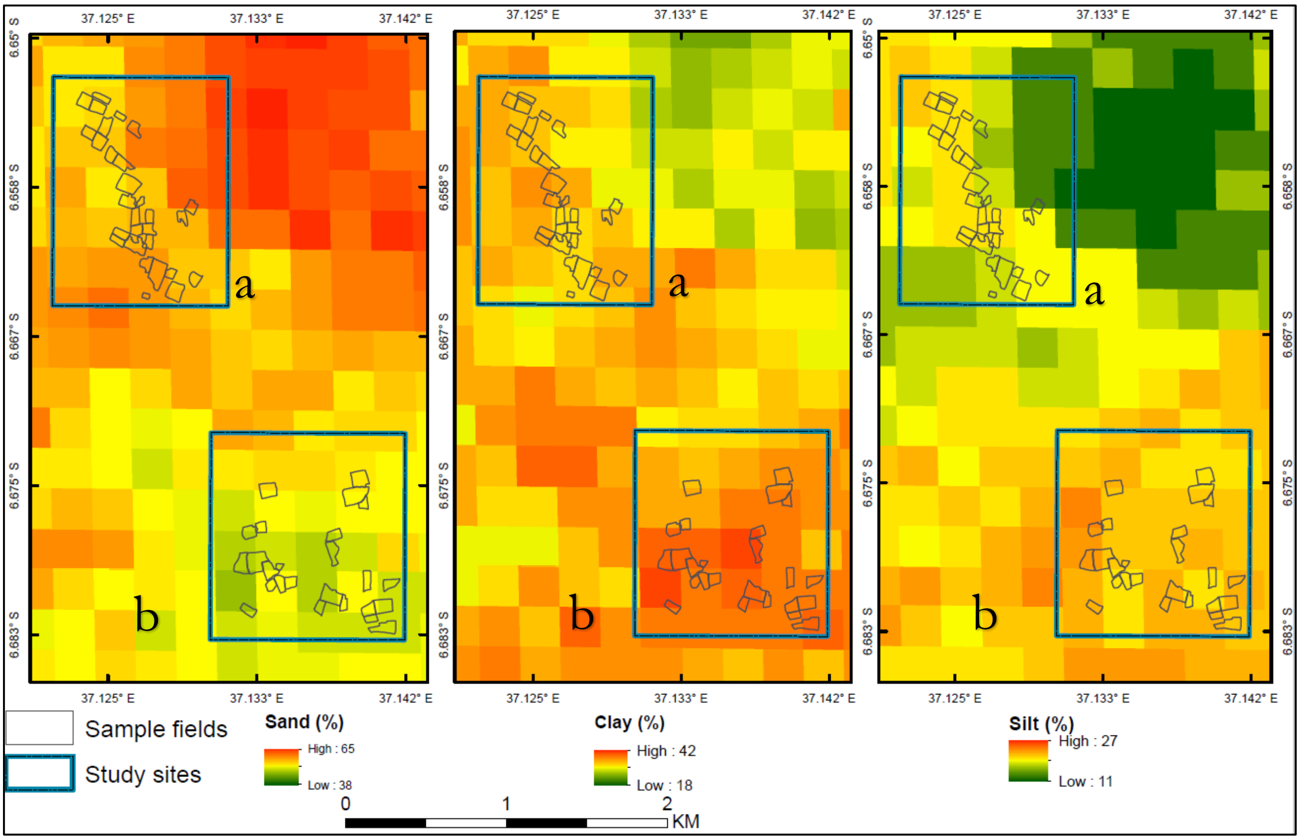


Figure 3: Spatial variation of percentage of soil texture in Gongoni (a) and Mbuyuni (b).The grey lines indicate the sampled maize fields.

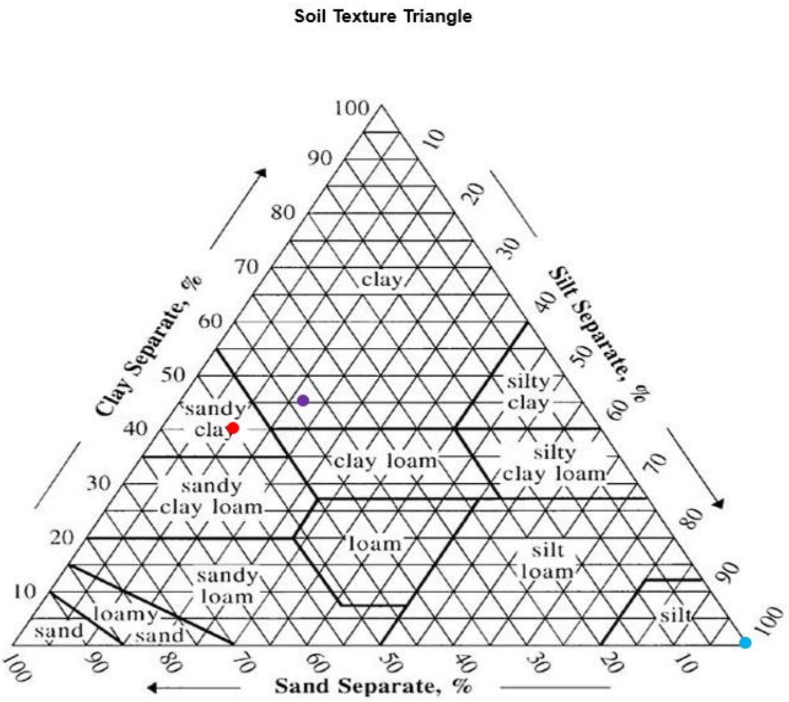
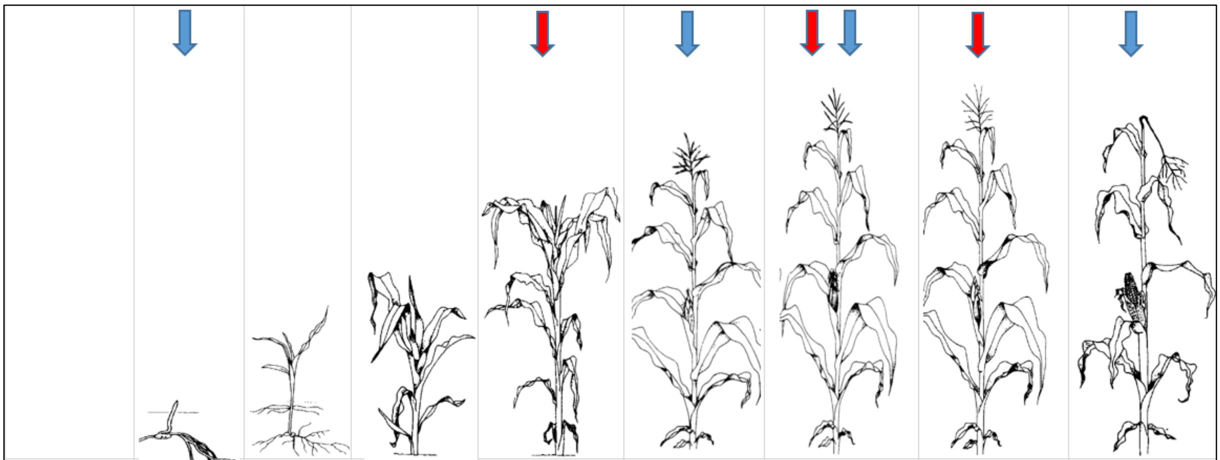


Figure 4: USDA Soil texture triangle showing proportions of - (a) Sandy clay in Gongoni indicated by red pointer and (b) Clay soil in Mbuyuni marked by purple pointer

More than 80 percent of the Kilosa population depends mainly on agriculture as a source of food and income. A variety of crops is grown on the two study sites which include maize, rice, millet, cassava, beans, and cowpeas. Apart from food crops, main cash crops include cashew nuts, coconuts, bananas and sugar cane. Small scale farming where the average farmland is less than one hectare represents 90 per cent of agriculture with large scale farming representing the remaining 10 percent (Kajembe et al., 2013). The small-scale farm holders are mostly subsistence farmers who produce mainly for domestic use, selling only their surplus to the nearby local markets. There is limited usage of inputs such as inorganic fertilizer, organic fertilizer or improved seeds with almost 95 per cent using hand hoes for cultivation.

The land ownership in Gongoni is leasehold as it was initially state-owned sisal plantation until 2000 when it was leased to the farmers, most who have cultivated for less than five years. In Mbuyuni, most land is family owned mainly inherited from grandparents with farming having been practiced in these fields since 1960's. The planting season coincides with the start of rainfall season in late February and early March. The farming system includes intercropping, mixed and mono-cropping. In most cases, the planting dates for the main crop and the intercrop has a span of two to three weeks which is different for mixed cropping system in which all the crops are planted at the same time. Common crops mixed with maize include pigeon peas, sesame, and cowpeas while intercrops include groundnuts and sunflower. Figure 5 indicates maize development stages and different transition period as described by (Meier, 2001; Ransom, 2013). The blue and red arrows indicate period in which WV and UAV imagery were acquired respectively.



Maize development	Germination	Leaf development	Stem Elongation	emergence, heading	Flowering, anthesis	Silking	Fruit Development	Ripening/ Senescence
(Meier, 2001)	Stage 0	Stage1,2	Stage3,4	Stage5	Stage6a	Stage6b	Stage7	Stage8,9
Field dates	01-Mar	14-Mar	28-Mar	18-Apr	02-May	16-May	06-Jun	04-Jul
(Ransom, 2013)	7 days	14 days	28 days	49 days	63 days	75 days	98 days	126 days
Imagery	14-Feb(WV2)			19-Apri(UAV)	13-May(UAV)	13 June (UAV/WV)	26-Jun(UAV/WV3)	22JUL (WV2)

Figure 5: Maize development stages with corresponding remotely sensed images and transition dates from one stage to another

2.2. Unmanned Aerial Vehicle (UAV) Data

Three fine resolution multi-spectral UAV and four WorldView-2 and -3 images were used in this study. The images were acquired on different dates during the maize growing season in the year 2015 as indicated in Table 1.

Table 1: Aerial and satellite imagery acquisition periods

Source	14 Feb 15	19 April 15	13 May 15	13 Jun 15	26 Jun 15	22 Jul 15
Sensor	WV- 3			WV-2	WV-3	WV-3
UAV RGB		UAV-Gongoni	UAV-Mbuyuni	UAV-Gongoni		
UAV NIR		UAV-Mbuyuni	UAV-Gongoni	UAV-Gongoni		
			UAV-Mbuyuni	UAV-Mbuyuni		

The UAV images were acquired using two cameras, Red-Green-Blue (RGB) and Red-Green-Near infrared (NIR) which were flown twice, each time with a different camera as shown in Figure 6. Both cameras had different spectral ranges as indicated in Table 2. The UAV-NIR camera was a modification of the original RGB camera using a band-pass filter to allow it detect radiation in NIR band (Lebourgeois et al., 2008). During the modification, the blue band was replaced with NIR band. The UAV carried on board a Canon S100NIR NIR camera with 12 megapixels controlled by the drone's autopilot.

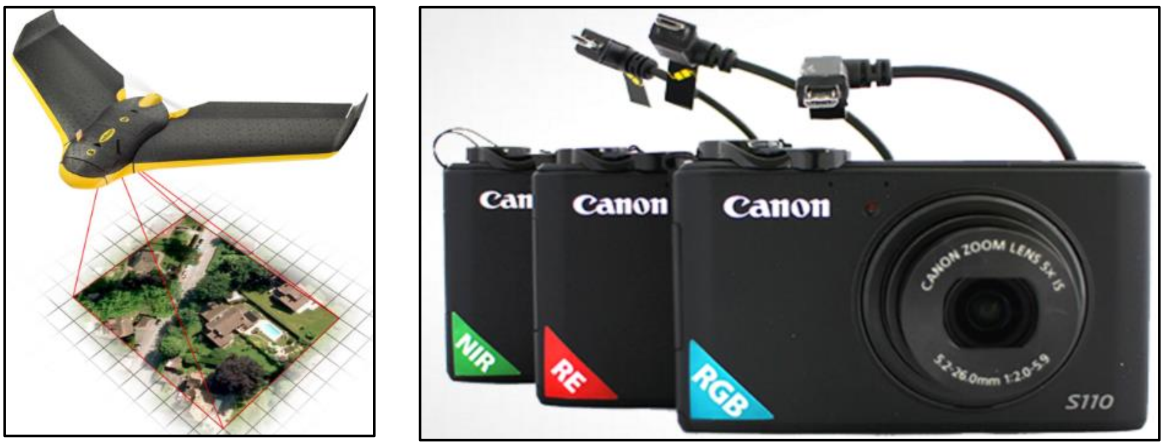


Figure 6: A fixed wing eBee UAV with different types of camera (Source: SenseFly, 2015)

The RGB and NIR wavelength response function for each the RGB and NIR sensors is indicated Figure 7

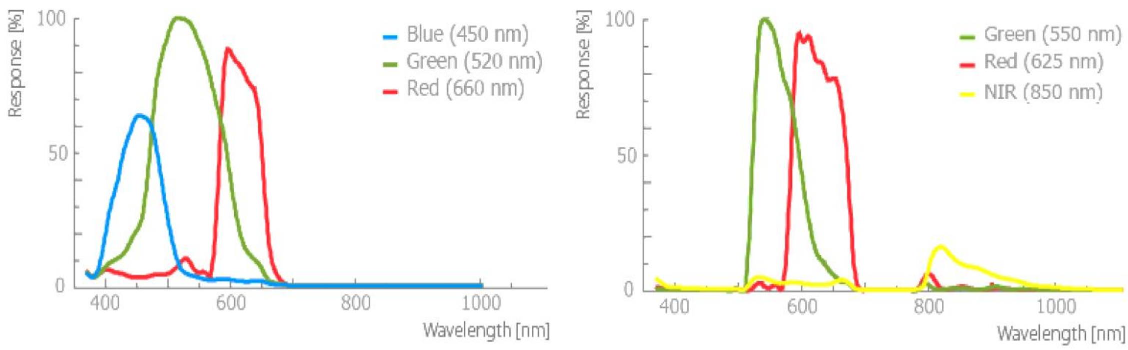


Figure 7: S100 RGB and NIR camera wavelength response function (Source: (Arellano, 2015)

Table 2: WV and UAV spectral band wavelengths

Satellite				**Aerial**			
Sensor	*Band*	*Wavelength range (nm)*	*Spatial reso.(m)*	*Image*	*Band*	*Wavelength range (nm)*	*Spatial reso. (m)*
WV-2				eBee			
	B2: Blue	450-510	1.6	RGB	B1:Red	575-725nm	0.05
	B3: Green	510-580	1.6		B2:Green	400-640nm	0.05
	B5: Red	630-690	1.6		B3:Blue	390-510nm	0.05
	B6: Red Edge	705-745	1.6				
	B7: NIR1	770-895	1.6				
	B8: NIR2	860-1040	1.6				
WV-3	B2: Blue	450-510	1.2	eBee NIR	B1:Red	575-675nm	0.05
	B3: Green	510-580	1.2		B2:Green	450-650nm	0.05
	B5: Red	630-690	1.2		B3:NIR	800-900nm	0.05
	B6: Red Edge	705-745	1.2				
	B7: NIR1	770-895	1.2				
	B8: NIR2	860-1040	1.2				

The UAV aerial imagery with a ground pixel resolution of 0.05 m per pixel at 114 meters above the ground surface was acquired using eBee Unmanned Aerial Vehicle (UAV), manufactured by Sensefly Ltd (Cheseaux-Lausanne, Switzerland). Field campaigns, flight planning, and actual imagery acquisition was carried by University of Maryland (UMD), USA in collaboration with Sokoine University of Agriculture (SUA) in Morogoro, Tanzania under the umbrella of STARS project. UAV imagery acquisition within the two 1x1 km study site in Kilosa (Figure 1) was carried out once every month beginning from April to June, which was the main maize farming season. Despite the initial idea of flying twice per month, it was decided to fly once a month due to field logistic challenges. During data pre-processing, there was a failure in generating RGB composites for Mbuyuni and NIR for Gongoni acquired on 19 March 2015. UAV image pre-processing was carried out by STARS project partners at the University of Maryland (UMD). The eBee has an inbuilt GPS unit that collects its position and an inertial navigation system that collects the camera orientation and angular parameters that are both necessary for proper image projection (Sharma et al., 2014). Orthorecfication was implemented automatically using eBee's Postflight Terra 3D software package. Radiometric calibration was carried out to convert digital numbers (DN) to the top of atmosphere (TOA) reflectance values. In order to reduce the effect of sun angle, data collection was scheduled between 10 am and 12 noon before the overhead sun and in a cloud-free atmosphere as suggested by Honkavaara et al., (2013). In addition, the atmospheric correction was not carried out as there was a minimal atmospheric effect due to low flying height (110 meter above the ground surface).

2.3. WorldView Data

WorldView-2 and WorldView-3 imagery were acquired by the STARS project from Digital Globe Company, an American commercial vendor of space imagery and geospatial content based in Longmont, Colorado, United States. The WorldView-2 and World View-3 imagery spatial resolution was 1.6 meters and 1.2 meters respectively. Total of 6 images were acquired between February 2015 and July 2015 which is the main maize growing period. However, two images for May and July had clouds and were not used in the study. Although most farmers planted in March, there were few others who planted in February and, therefore, the 14 February image qualified to be used.

The sequence of correcting WV images was first radiometric calibration which involved the conversion of digital numbers (DN) to the top of atmosphere reflectance using the physical gain parameters contained in the satellite metadata file. Thereafter, the atmospheric correction was carried out using Second Simulation of a Satellite Signal in a the Solar Spectrum Vector (6s) radiative transfer model specifically adjusted for the Digital Globe data which includes WV imagery (Vermote et al., 2006). The algorithm uses external information derived from MODIS for aerosol and atmospheric condition estimation on the day of image acquisition to correct the effect of aerosol and a gaseous particle that might have had an effect on the reflectance received by the sensors.

The last step was orthorectification process which was applied using satellite-derived geometric metadata. The pre-processing procedure of WV images was carried out by STARS project team here in ITC. Although the geometric correction was carried out with high precision using automated workflow for both datasets, there was location shift between UAV and WV which was manually corrected by editing WV image header file so as to shift X and Y pixel location to a point where features such as roads and buildings showed an almost a perfect merge with UAV imagery. The location shift of the WV image pixels shifted results are summarized in Table 3 and Figure 8 shows the flow chart describing the pre-processing steps used on WV data.

Table 3: WorldView geometric shift to fit UAV imagery

Satellite_ Image ID	Date	Spatial Resolution (m)	X shift (pixels)	Y shift (pixels)
54330600010	14-02-15	1.2	148	-54
54460880010	13-06-15	1.6	42	48
54487783010	26-06-15	1.2	-139.5	13
54551817010	22-07-15	1.6	48	-54

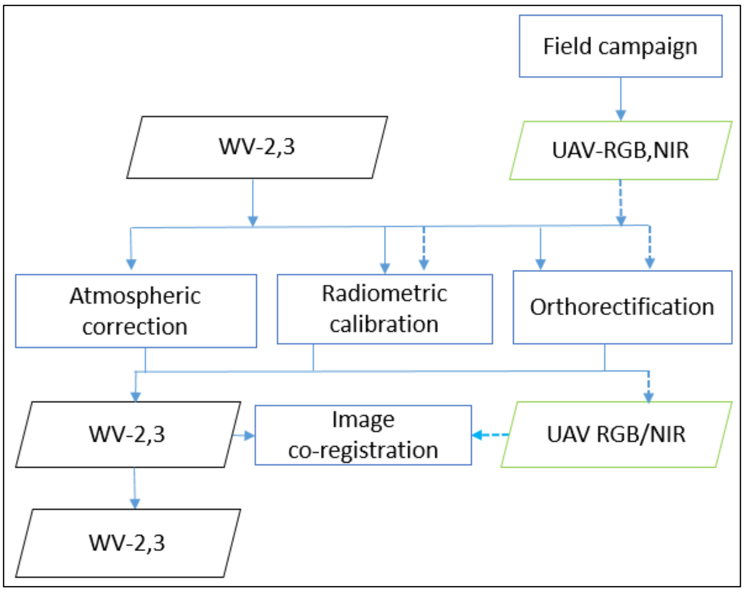


Figure 8: Flowchart describing the WV and UAV pre-processing steps

2.4. Integration of same date WorldView and UAV imagery

The availability of UAV and WV of the same date (13 June 2015) provided an opportunity for creating WV VI of 13 May 2015 using a regression model to establish the relationship between reflectance of UAV and WV imagery. Integration used in this context meant using relationship established from two remote sensing data acquired by different sensors to generate a new image of a different date if imagery of one sensor is available. This was carried out to determine if possible to integrate aerial and satellite data so as to fill data availability gap due to limitation such as of cloud on WV images and in case there is a failure in acquiring UAV image, then an alternative approach is available. First a test was carried out by correlating the average field-level NDVI between the two datasets which showed a good relationship with R^2 =0.77. The result indicated that the two images responded almost similarly to vegetation reflectance and, therefore, an attempt was made to integrate the images using simple linear regression model.

The first step was to compute the NDVI of the two images (WV and UAV) at their original spatial resolutions. The second step was to resample both the WV and UAV to 8 m spatial resolution using nearest neighbourhood technique. The reason for resampling to a coarser resolution was to ensure a complete overlap of the pixels so that a linear relationship could be established between NDVI reflectance's of the two sensors. The regression equation would help in determining the reflectance bias within the two sensors which would then be applied to an NDVI image of different date (either WV or UAV). The computation of VI was to harmonize the differences in band reflectance from the two images. To compute regression equation, pixel values were randomly selected from 780 points and coefficient of determination computed. A square buffer of 0.8 meters was generated and mean NDVI values computed using zonal statistics for each of the random points. The average NDVI within each of the 780 randomly selected points were exported and coefficient of determination computed which gave an R^2 =0.51 (n=780). The reason field-averaged NDVI was not used was to minimize pixel contamination and therefore choice of small area was preferred. It was therefore assumed that there was minimal heterogeneity within the 0.8m square buffer

$$Y=0.6577x + 0.1343 \quad (1)$$

Where Y is the VI values of the new WV generated; x is the UAV pixel values acquired on 13 May 2015 and 0.134 is the reflectance bias error. The regression equation was applied to the UAV imagery of 13 May 2015, taking its pixels values as the independent variable and WV as the dependent variable so as to compute a WorldView NDVI map of 13 May 2015 at 8 meter spatial resolution.

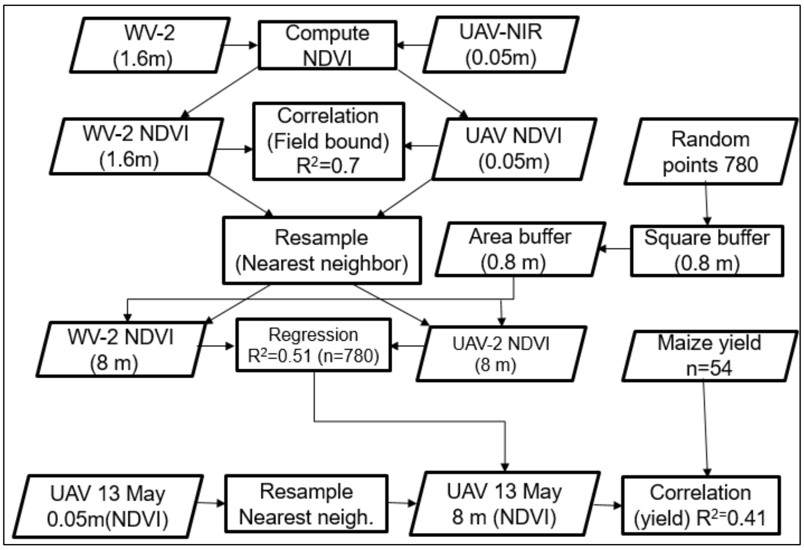


Figure 9: Flow chart showing the steps taken when computing the integrated imagery NDVI from UAV

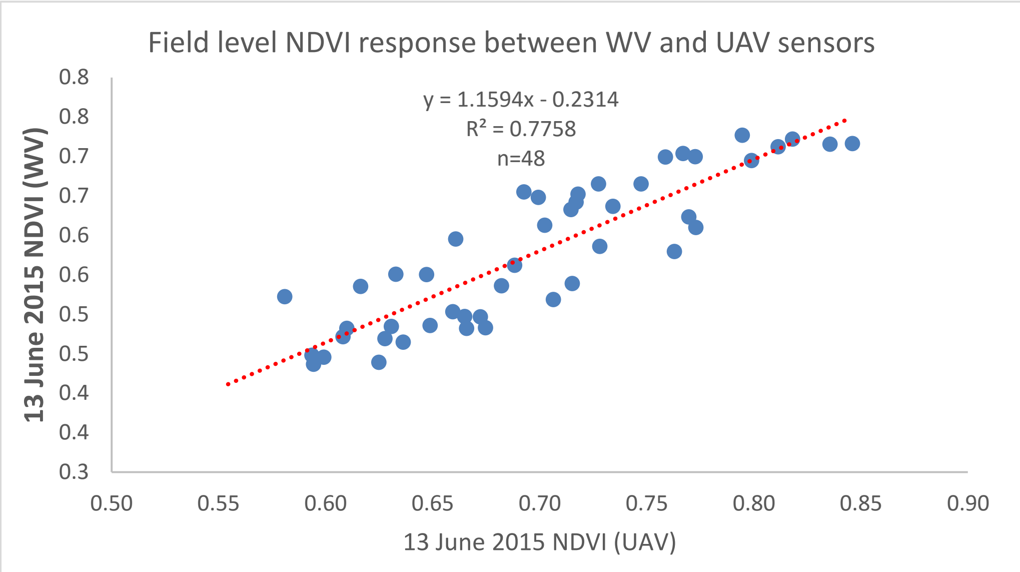


Figure 10: Scatter plot of field level NDVI between same data UAV and WV imagery

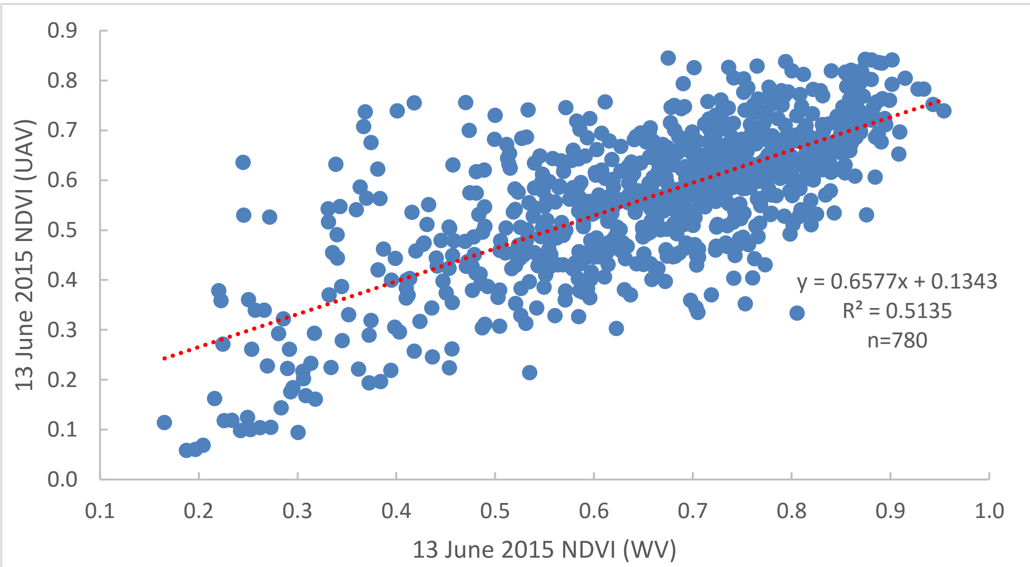


Figure 11: Equation applied to derived new WV NDVI imaged from same date UAV and WV NDVI VI imagery

2.5. Field-level interview data

The purpose of the fieldwork was to collect data on maize production, harvested area, and management activities carried out at field-level during the maize growing period between February and July 2015. The field work was conducted from 28th September and ended on 16th October 2015. The farmer interview was carried at the location where maize was grown for two main reasons, first for the farmer to show the extent of his field thus ensure accurate field delineation and secondly to collect field-based location data such as the observed difference in yield within the field. Total of 54 farmers were selected using a purposive random sampling approach in which twenty-eight (28) farmers drawn from Gongoni and twenty-six (26) from Mbuyuni.

Purposive and random selection of the interviewed farmers was carried out based on whether their fields were part of the farm management units (FMU) monitored by the STARS project team during the entire maize growing season, secondly the estimated maize production as perceived by the farmers on a scale of high-average-low production. The reason for asking the farmer to identify high-average-low production fields was to capture a large range of occurring yield levels which provide representativeness of the un-sampled fields and help in explaining maize yield difference within and between fields. To further categorize the fields in terms of high-medium-low production range, visual check on the UAV image acquired on 15th May 2015 gave an idea of the maize status since it was possible to identify farms with green, pale green and yellow colored section of the fields. Maize production was reported in a number of bags of maize cobs per field. To convert this to standard units, two bags of maize cobs contained in standard large sized gunny bags were converted to a one-100kg bag of shelled maize. The approximation was reached upon after wide consultation with farmers and local agriculture extension officer.

Additionally, field management practices such as date of sowing, harvesting, weed and pesticide control, cropping system, land ownership, source of seeds, period of planting (whether before or after rains), tilling method applied, fertilizer/manure applied shocks experienced during the growing season, The farmer response during the interview was keyed into CSEntry Android programmed App downloaded from Google Play Store using android phone. Every new entry was captured as record and automatically assigned an ID which was entered separately in the table with unique ID and timestamp and the end of the day, it was downloaded and errors checked and corrected before the leaving for the field the next day. The advantage of CSEntry as compared to paper based interviewed is that it reduced data entry errors and time. In addition, it allowed collection of geo-tagged photos which facilitated post-field data analysis.

The maize fields were digitized using the 13th May 2015 UAV-RGB imagery as the background layer. The image was chosen given the difference between maize and non-maize fields was distinct. It is also important to note that the GPS points collected around the field with the guidance of the farmer were overlaid on the image so as to show the exact extent of the field section where maize was grown. To ensure the digitized fields merge with field production and management interview data collected using the tablet, the same unique identifier was used for each digitized field and the corresponding household interview data. Considering the effect of trees and vegetation along the field boundaries edges on the computation of field-level image band mean, the original field polygon was shrunk by 70% of the original area using QGIS vector buffer by percentage plug-in to generate the boundaries shown in Figure 12. Maize field area was computed using field calculator in ArcGIS software. The maize field area was computed based on the original digitized boundary and not the 70% shrunk boundary.

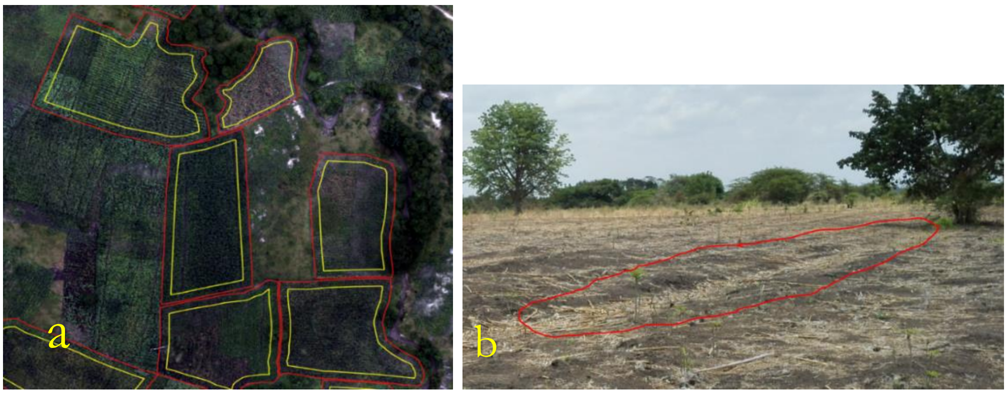


Figure 12: Digitized maize field boundaries(red) with 70% shrunk field boundaries (yellow); (b) photo taken during the field work showing the fuzzy boundary between two adjacent maize fields separated by a tree.

3. METHODS

3.1. Spectral indices and metrics computation

Spectral vegetation indices were calculated based on the UAV-RGB, UAV-NIR, and WV images. The mean spectral bands values were extracted for each of the UAV and WV image using zonal statistics in QGIS. The shapefile with unique field identifier, farmers name, field area, maize production and computed maize yield (ton/ha) were appended to the corresponding mean band value. This was performed separately for the WV and UAV imagery. The computed vegetation indices are listed in Table 4. The choice of the vegetation index was considered based on index that has been most applied in maize yield modelling, vegetation index that considers the use of RGB band section of electromagnetic spectrum and which computation algorithm applied includes either ratio or band difference.

Table 4: Vegetation Indices evaluated in the study

Vegetation Index	*Equation*	*Author*
Normalized difference vegetation index	**NDVI**=(NIR−R)/(NIR+ R)	(Rouse et al., 1974)
Modified soil adjusted vegetation index	**MSAVI**=0.5{2*NIR+1−√[(2*NIR+1)2−8(NIR−R)]}	(Qi et al., 1994)
Enhanced vegetation index	**EVI**=2.5(NIR − R)/(NIR + 6*R − 7.5*B + 1)	(Huete et al., 2002)
Visible atmospherically resistant index	**VARI**=(G−R)/(G+ R−B)	(Gitelson et al., 2002)
Difference Vegetation Index	**DVI**=NIR-R	(Tucker, C. 1979)
Transformed chlorophyll absorption reflectance index	**TCARI**=3[(Redge−R)− 0.2(Redge−G)(Redge/R)]	(Haboudane et al., 2002)
Excess Green Index	**ExG**=2*G-R-B	(Woebbecke et al., 1995)
Green Difference Vegetation Index	**GDVI**=NIR-G	
Green Normalized Difference Vegetation Index	**GNDVI**=(NIR-G)/NIR+G	(Gitelson et al., 1996)
Ratio vegetation index (also called simple ratio)	**RVI**=NIR/R	(Jordan C.F., 1969)
Green leaf index	**GLI**=(2*G − R − B)/(2·G + R + B)	(Louhaichi et al. 2001)
Green Atmospheric Resistant Index	**GARI**=NIR-[G-1(B-R)]/NIR+[G-1(B-R)] where	(Gitelson et al., 1996)
Green Ration Vegetation Index	**GRVI**=NIR/G	(Sripada, R. et al 2006)

3.2. Maize yields modeling approach

To estimate maize yield from the WV and UAV images, linear relationships were established between field-level maize yield data obtained from the interviews and field-averaged vegetation indices. A variety of vegetation indices were computed from each image and vegetation indices metrics were used to develop different models. Adjusted coefficient of determination, Root means Square Error, bias-corrected and accelerated confidence interval were the parameters used to assess model performance computing using the bootstrap resampling technique in R. The following section will explain in detail step by step process that was undertaken to derive maize yield relationships models and the criteria for selecting the optimal one.

Statistical analysis was carried out on the field-level yield interview data to determine if the data were normally distributed. According to Shapiro-Wilk test (W=0.000, p=0.05) (Figure 13) the data were not normally distributed. This has an effect in computing statistical analysis such as regression and analysis of variance at it will lead to bias (the result being not be representative of the population). To overcome the normality issues with the data, a simple linear regression model was used to calibrate the model and bootstrap resampling technique applied for validating the model. Although there are many other available statistical validation techniques which have been applied to validate VI-yield relationships such as cross-validation and Jack-knife method, bootstrap resampling was preferred due to non-normal nature of the data and the small sample (n=54) to allow application of either cross-validation or jack-knife method. Furthermore, bootstrap method is less biased and with less coefficient of variation as compared to Jack-knife and cross-validation methods (Efron et al., 1983). However it is limited in that it relies on a representative sample and has got high variability as a result of finite replication commonly referred to as Monte Carlo error (Koehler et al., 2009)

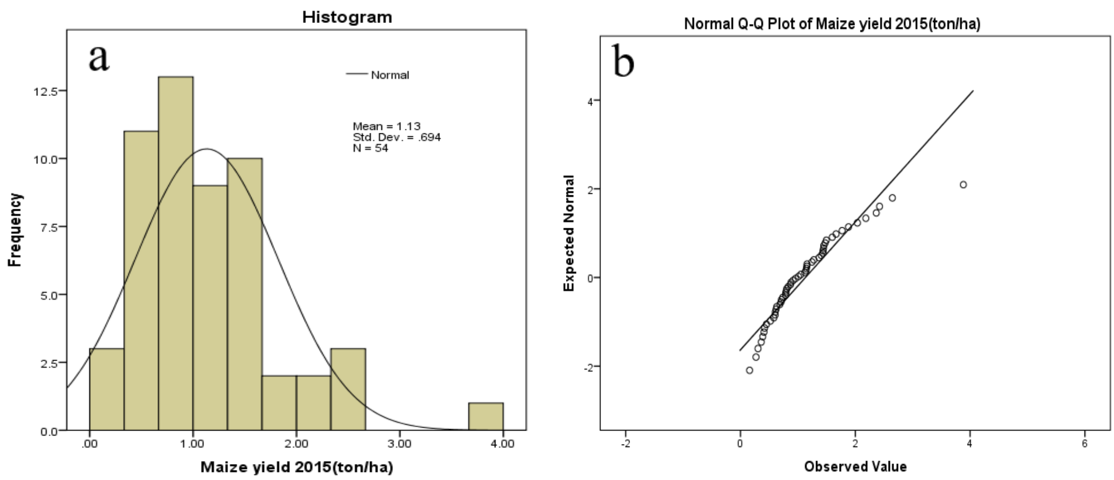


Figure 13: Non-normal distributed maize yield data (a) bar graph fitted with normal line and (b) Q-Q plot which shows the deviation of the distribution within the normal fit line.

The bootstrap, like cross-validation, is a data resampling approach (same data used several times) in order to derive mean, standard error of prediction and bias-corrected and accelerated (BCa) interval. It is a resampling method with replacement from the target population and this means the sample drawn may have some of the data represented several times. As a rule of thumb, the sample should be more than the square of the samples (in this case 54*54).

$$Yieldmodel < -lm(x \sim y, data = d) \quad (2)$$

Where yield model (Equation 2) is described as a function of x; which is the dependent variable (in this case maize yield); y the independent variable (VI metrics) and d; sampled field-level yield data The model summary statistics provides coefficient for the model (the constant and the slope) R^2, Adjusted R^2 and level of statistical significance (p-value). Once the model is derived, the next step is to validate the model using the bootstrap resampling method (Equation 3).

$$Yieldmodel.boot < -lm.boot(Yieldmodel, R = N) \tag{3}$$

In this case $Yieldmodel.boot$ is a function of the relationship established in the liner equation (2) applied with resampling (R) for a number of times N (approximately 54 * 54~3600). The summary of the model provides the model coefficient and the validated R^2 which in this study, the model with the highest validated R^2 is selected as the optimal model. An example of the function as applied in R-software is presented in the screen dump in (Figure 14) for the maxGARI derived from WV imagery.

```
###Scrip description############
#The R-Script fits the data into a simple linear regression equation
#It then computes the R-Squared and adjusted R-Squared of the model
#Computes model reliabilit by resampling n*n using Bootstrap resampling methods-method to validate the model
#Results are R-Squared, bias, standard error and bias-corrected and accelerated (BCa) interval(95%)
#Script source:Steve Kibet @2016
#Adapted from non-parametric Boostraping (http://www.statmethods.net/advstats/bootstrapping.html)

#Load the data###############

setwd("c:\\Rprogramming")

c<-read.csv("Maize yield VI WV.csv");
names(c)

#####Obtain R-Squared and adjusted R-Squared from the data
library(simpleboot)
MaxGARI<-lm (Maizeyield~MaxGARI, data= c)#(fits the x,y data into a simple linear regression model)

MaxGARI.boot <- lm.boot(MaxGARI, R = 3600)#(Boostrap resample 3600 with replacement)

summary(MaxGARI)#()
summary(MaxGARI.boot)

# function to obtain R-Squared from the data
library(boot)
rsq <- function(formula, data, indices) {
  d <- data[indices,] # allows boot to select sample
  fit <- lm(formula, data=d)
  return(summary(fit)$adj.r.squared)
}
# bootstrapping with 3600 replications (n=54; 54*54=2916~3600)
results <- boot(data=c, statistic=rsq,
                R=3600, formula=Maizeyield~MaxGARI)

# view results
MaxGARI #(model coefficinets)
print(results) #(prints the results of Bootstrap statistics model, the R-Squared, bias, and standard error )
plot(results)#(plots the histogram of t-test to determine normality of the data, it includes the Quartile plots)

# get 95% confidence interval-computes bias-corrected and accelerated (BCa)
boot.ci(results, type="bca")

####################End######################
```

Figure 14: Bootstrap scrip applied for validating VI-maize yield relationship

Besides evaluating the relationship between single-date VI and maize yield, a number of VI temporal integration approaches were used that combine VI information from multiple dates. These included maxVI and cumVI. These vegetation index metrics are important in studying vegetation development, for example, looking at its phenological characteristics such as germination, leaf emergence and the start of senescence (Vrieling et al., 2011). In addition, single-date VI would be incompatible with yield estimation equation since the simple regression would neglect man-induced factors which have an eventual effect on yield increase (Huang et al., 2013). Furthermore, longer VI integration periods minimizes variability in yield prediction as results of variations in image acquisition dates, processing and difference in management factors such as early or late planting. The VI metrics around period of maximum VI have shown to be strongly correlated to maize yield (Mkhabela et al., 2011). In most of the crop yield studies, periods around flowering and fruit development has shown to have high yield-reflectance relationship (Laigang Wang et al., 2014). On the other hand, VI changes outside this period (i.e. early and late in the season) have shown to have a poor relationship with yield. Therefore, most studies have concluded that the period between mid-late growing periods is a good indicator of yield. Considering the optimal period has been established around the period of maximum VI, MaxVI metrics was tested to determine if season's maximum VI can provide better yield estimate.

Therefore, the first approach was to extract mean band surface reflectance values using zonal statistics from the digitized field boundaries and exported to Microsoft excel for VI computation. The second step was to compute the various single-date VI described in Table 4. Studies have shown that the longer the VI integration, the minimal the variability in yield prediction (Laigang Wang et al., 2014). The maxVI was derived from the highest VI value derived from each single-date image which was assumed to be equal to the peak value of the seasonal VI. Data interpolation was not applied to single-date VI in order to determine the value at each single period of crop growth. Summary of computed metrics is presented in Table 5.

Table 5: Vegetation index variables and the calculation formulas

Vegetation index variables	Description of formulas
VIw_1	14th Feb. 2015 WV-VI index
VIw_2	13th May. 2015 WV-VI index
VIw_3	26th Jun. 2015 WV-VI index
VIw_4	22th Jul. 2015 WV-VI index
$cumVIw_1\text{-}w_2$	$VIw_1 + VIw_2$
$cumVIw_1\text{-}w_3$	$VIw_1 + VIw_2 + VIw_3$
$cumVIw_1\text{-}w_4$	$VIw_1 + VIw_2 + VIw_3 + VIw_4$
$cumVIw_2\text{-}w_3$	$VIw_2 + VIw_3$
$cumVIw_2\text{-}w_4$	$VIw_2 + VIw_3 + VIw_4$
$cumVIw_3\text{-}w_4$	$VIw_3 + VIw_4$
$maxVIw_1\text{-}w_4$	$Max(VIw_1 ;VIw_2 ;VIw_3 ;VIw_4)$
VIu_1	19th Apr. 2015 UAV-VI index (RGB & NIR)
VIu_2	13th May. 2015 UAV-VI index (RGB & NIR)
VIu_3	13th Jun. 2015 UAV-VI index (RGB & NIR)
$cumVIu_1\text{-}u_2$	$VIu1 + VIu_2$
$cumVIu_1\text{-}u_3$	$VIu_1 + VIu_2 + VIu_3$
$cumVIu_2\text{-}u_3$	$VIu_2 + VIu_3$
$maxVIu_1\text{-}u_4$	$Max(VIu_1; VIu_2; VIu_3)$

The maize yield estimation model were evaluated using the following indicators:-
Root mean square error (RMSE):

$$\mathrm{RMSE} = \sqrt{\sum_{i=1}^{n} \frac{(Y_i - Y'_i)^2}{n}} \qquad 4$$

Where, n is the number of observation Y_I is the observed and Y'_I is the predicted value

Adjusted coefficient of determination:

$$AdjR^2 = 1 - (1 - R^2)\left[\frac{n-1}{n-(k+1)}\right] \quad 5$$

Where n = number of sample and k the number of independent variable in the regression

Apart from yield data collected, management factors such as tilling methods, cropping system source of seeds planted, the frequency of weeding, sowing date, the level of pest and diseases infestation were analyzed to determine if it had a significant difference in maize yield. One way analysis of variance (ANOVA) approach was used to determine the influence of management factors on maize yield. The prediction accuracy of the different modeling strategies was assessed by Root Mean Square Error (RMSE). The test under the null hypothesis (H_o) was that there was no significant differences ($p > \alpha > 0.05$) between the different management factors applied at field-level on maize yield and an alternative hypothesis (H_1) that the management factors applied by the farmers had significant differences on maize yield ($p < \alpha < 0.05$). In the event the $p < \alpha$ (which indicates no difference between the groups) a further test between the combination of different groups was performed using Fisher Least Significance Difference (LSD) method using Equation 6 as described in (Williams et al., 2010). The rationale behind the LSD technique is that when the null hypothesis is true, the value of (t) statistics evaluating the difference between group's a_1 and a_2 is equal to zero.

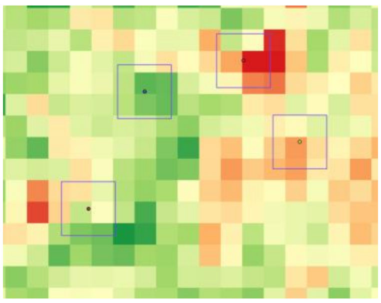
$$|M_{a1+}\ M_{a2}\ | > LSD = t_{v\alpha}\sqrt{MS_{s(A)}(\tfrac{1}{Sa1} + \tfrac{1}{Sa2})} \quad 6$$

3.3. Field-level maize yields spatial variability

The field level yield variability map was derived based on the optimal index which gave the highest adjusted correlation coefficient with actual maize yield with low RMSE and bias error. The selected model was applied to the best performing VI imagery to derive yield variability map. Since the VI derived from the different images was computed based on mean band values extracted, the optimal VI was computed so as to derive VI of each pixel and the model equation applied. The result was a yield map in which each pixel represent maize yield in that specific location in tons per hectare.

A further test on the effect of yield management factors influences on maize yield. Using the sample points ranked in order of high-average-low yield was tested using Spearman's rank correlation. This was achieved by first creating a square buffer of 4 meters around the sample points to gather for GPS errors and geometric correction errors. Secondly zonal statistics was applied to extract yield values corresponding to the-the farmer reported yield rank. This was then exported to Microsoft Excel and non-parametric Spearman's rank correlation analysed to determine to what level of accuracy are the location the farmer reported high yield correspond to high VI values. The Figure 15 shows some of the in-field sampled data with a 4-meter buffer used to aggregate yield data within those pixels.

Figure 15: Field-level reported maize yield rank locations with the 4-meter buffer



4. RESULTS

4.1. Correlation coefficients between maize yield and VI metrices tested

The coefficient of determination (R^2) between maize yield and the VI variables derived from UAV-RGB, UAV-NIR, and WV-NIR VI metrics are summarized in Figure 16.The colored cells indicate R^2 of maize yield-VI relationship at a given date. The point at which the same dates converge in both X and Y axis indicate R^2 derived from a single-date VI imagery while different date's combination indicates cumVI between the selected dates. The top section of the chart indicates the maximum VI-yield relatioship.

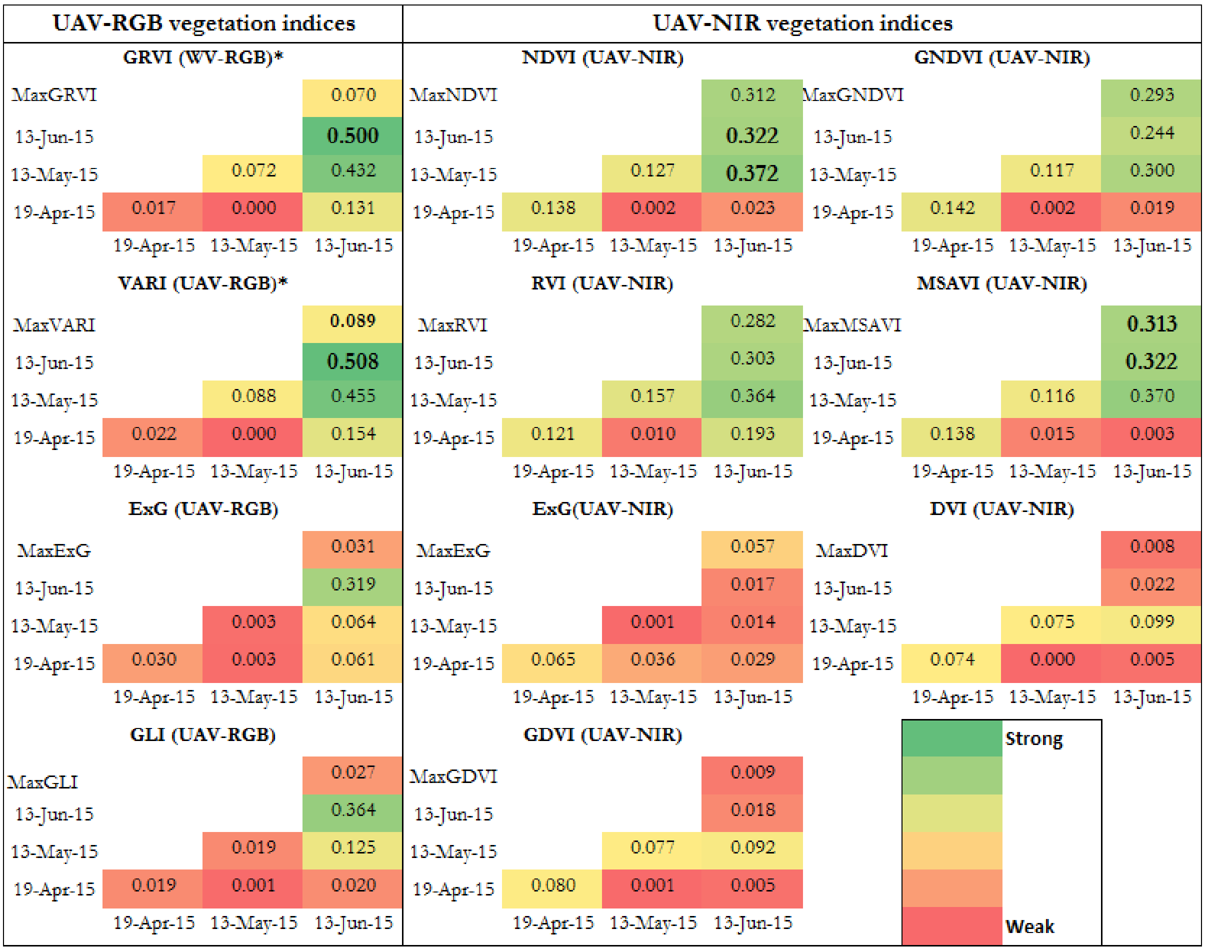


Figure 16: Coefficient of determination of maize yield and temporal VI metrics derived from UAV-RGB and UAV-NIR images. For the majority of the fields, the dates 19th April 2015 correspond to inflorescence stage of maize stage, 13th May 2015 flowering, and 13th June 2015 silking stage. The R^2 was significant at the $p < 0.001$ except for the lowest $R^2 < 0.02$.

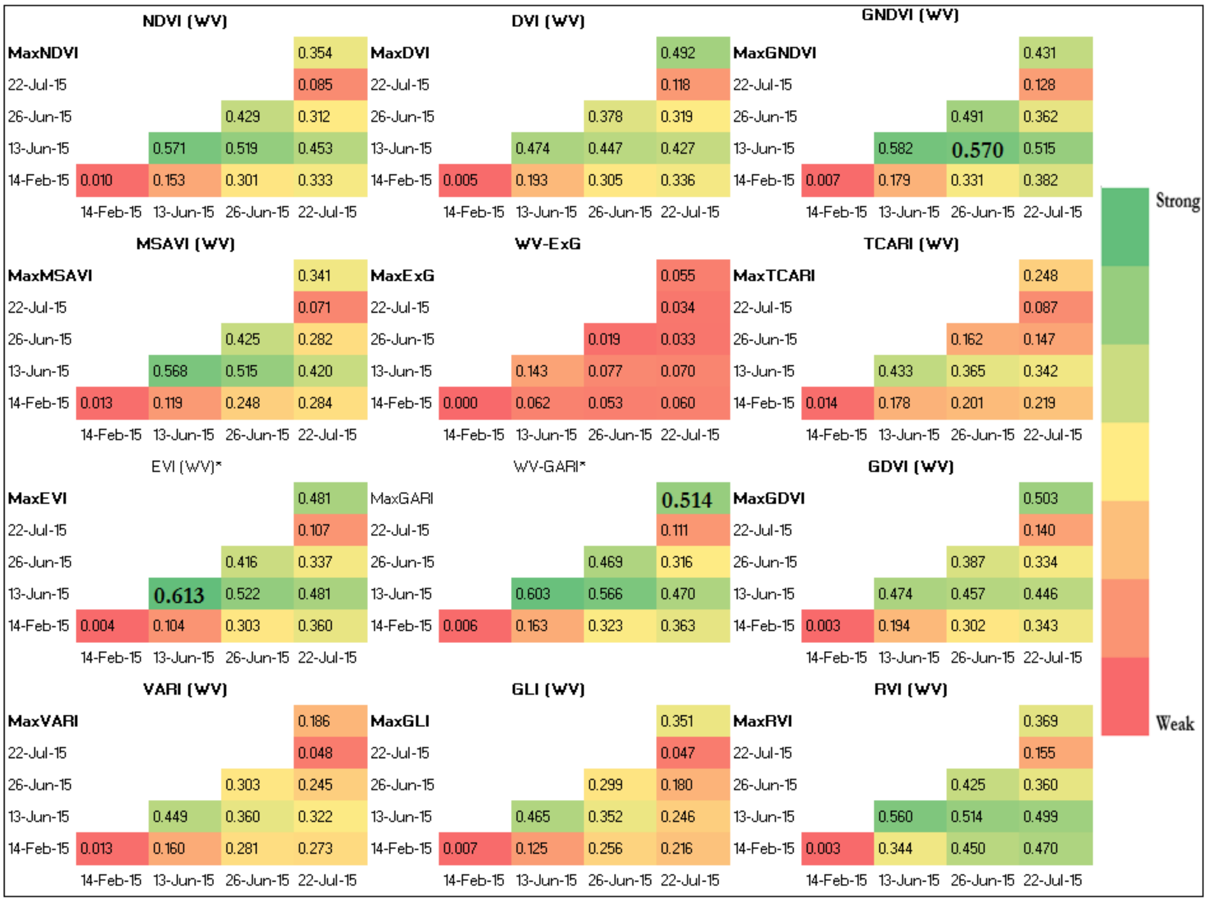


Figure 17: Correlation coefficient of maize yield and WV-VI data and maize yield at different stages of maize development. In most fields 14th February 2015 correspond to sowing period; 13th June 2015-silking, 26th June 2015-fruit development and 22nd July 2015 senescence period. The R^2 has at $p < 0.001$ except for the lowest $R^2 < 0$.

The correlation coefficient between maize yield-VI metrics obtained from single-date images indicates a weak maize yield-VI relationship at the beginning of the maize growing period, whereas the relationship between for imagery during the silking stage is significant for both the UAV and WV vegetation indices. The largest correlation coefficient derived from VI metrics and yield computed from UAV-RGB imagery was GRVI (R^2=0.5) and VARI (R^2=0.508) while for UAV-NIR was NDVI (R^2=0.322) and MSAVI (R^2=0.322) and with regard to WV-NIR, EVI (R^2=0.613) and GARI (R^2=0.603. The possible reason could be that around this period active photosynthetic activity is taking place in the maize and, therefore, any interference during this period such as insufficient water supply, nutrients, and disease or pest infestation would have adverse effects on maize yield. One notable observation is the weak relationship of maize yield-VI in the month of February whereby there was little vegetation including weeds considering the fields had been sowed.

Furthermore, the maize-VI relationship seems to deteriorate after a period of maximum greenness which is estimated around 13 May 2015 given most maize fields had consistently high VI values around this period. This could be as a result of declining photosynthetic activity as a result of a reduction in chlorophyll content which NIR vegetation index is most sensitive to. Therefore, the reflectance reaching the sensors is reduced as the maize heads toward senescence period. However, the WV imagery acquired during the senescence (22 July 2015) period had a stronger relationship with yield as compared to imagery acquired in the month of February. This was expected considering there were fields with mixed crops such as pigeon peas, sunflower and also fields planted late were still green. Although at inflorescence stage the maize crop is fully grown and characterized with maximum greenness, it still showed a small R^2 value (<0.1) with UAV imagery. This can be attributed to differences in weeding whereby in some fields, weeding had been completed while in some it was in progress. In addition, some fields though weeded had good maize crop but had lower VI because of weed removal. Apart from this, other vegetation growing in the maize field (either intercropped or mixed) contributed to high VI value considering they were almost same height as maize plant. Figure 18 provides insights on the status of the maize field during this period.

Figure 18: Very fine-resolution RGB image acquired on 13 May 2015 (flowering stage) showing maize field with (a) mixed sunflower with same height as maize (b) half weeded maize field (c) mono-cropped maize yield at inflorescence stage and (d) Mono cropped maize field with patches of weeds at flowering stage.

As maize crop grows towards silking stage (one month later), the relationship is seen to have improved with VARI derived from UAV-RGB and EVI derived from WV-NIR large R^2. Surprisingly, though, increased yield (highlighted in red in scatter in Figure 19) shows to correspond to decreased VI. This is contrary to what the model describes i.e. increase in VI corresponds to increased yield. A plausible argument for this unusual pattern could be the accuracy of interview data collected. There is the possibility of farmers having over reported maize production which leads to such inconsistencies whereby the yield does not correspond to the VI values. Secondly, it could be variation in planting dates which leads to differences in stages of maize development. The other reason could be related to single-date imagery VI which provides reflectance of a single period which is varies from field to field.

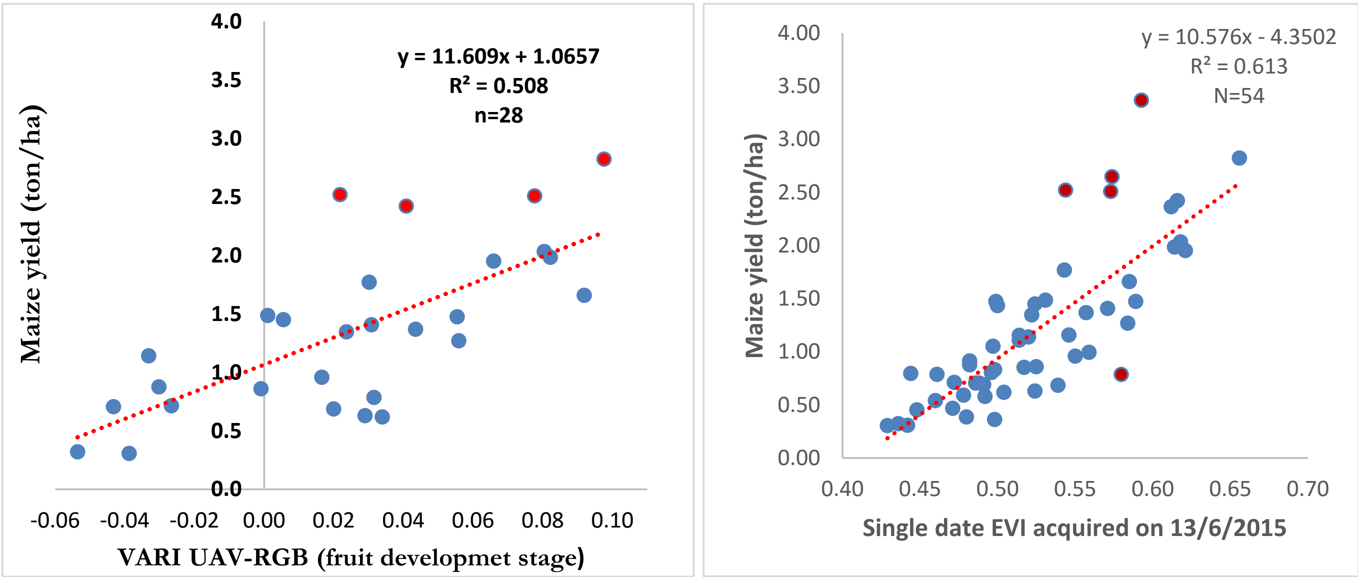


Figure 19: Scatter plot showing the relationship of maize yield with single-date (a) VARI (UAV-RGB) and (b) WV-EVI during silking maize growth stage. The red points indicate unusual pattern of yield which corresponds to low yield.

The cumulative UAV-RGB and UAV-NIR vegetation index results indicate weak VI-maize yield relationship the largest being cumVARI (R^2=0.455) during the flowering and fruit development as compared to VI derived during silking and fruit development cumNDVI (R^2=0.372). As compared to WV derived index cumGNDVI (R^2=0.570), showed better maize yield-VI relationship during silking and fruit development. This shows that cumVI derived from WV performed better than UAV-RGB and UAV-NIR cumVI's which could have been majorly contributed by the difference in image acquisition dates. As noted in the results, the inclusion of a longer VI integration period result to the weak maize-VI relationship as compared to shorter integration period from the flowering period. The other factor (though not directly tested) could be as a result of narrow WV spectral range as compared to broader UAV spectral range which might have resulted in a difference in sensor sensitivity to maize vegetation reflectance. The observed variation is better explained by the scatter plots in Figure 20.

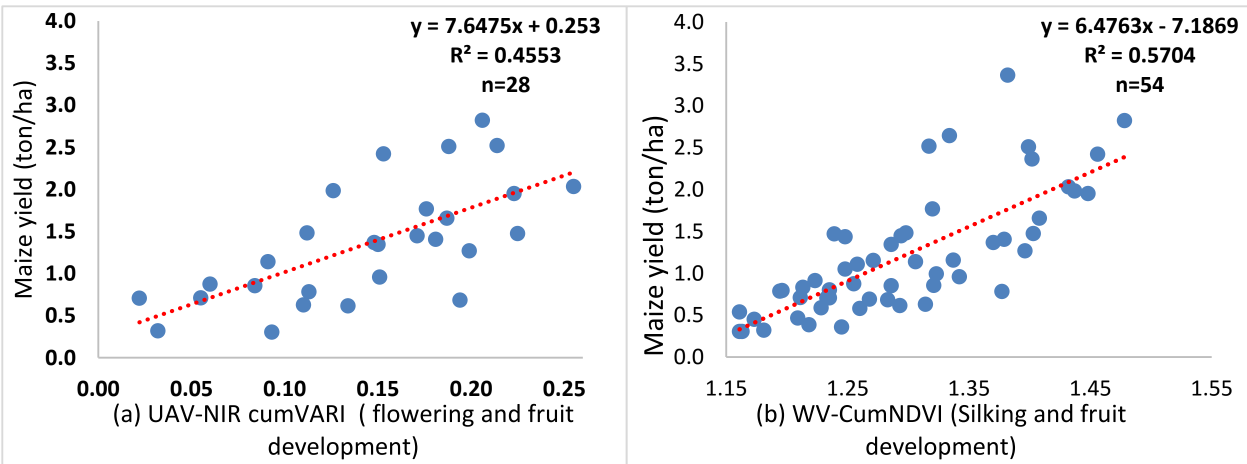


Figure 20: Maize yield-cumVARI relationship derived from UAV-NIR during flowering-fruit development stage and (b) WorldView Maize yield-cumNDVI relationship during silking-fruit

Maximum VI showed consistently small R^2 values in all the VI's tested in both the UAV-RGB and UAV-NIR sensor. The best UAV-RGB result was derived from maxVARI (R^2=0.089) while for UAV-NIR was maxMSAVI (R^2=0.313). In regard to WV, maxGARI (R^2=0.514) gave the best relationship with maize yield. One probable reason for the variable performance of maxVI is the confounding effect of weeds and other crops (sunflower and pigeon peas) grown together with maize in the same field which gave a high VI value. The performance of season's maxVI derived from WV was higher than what was observed with RGB derived VI's and this could be as a result of the difference in dates of image acquisition.
Scatter plot of the yield-VI variation derived from maxMSAVI and maxGARI during the maize growing season is indicated in Figure 21.

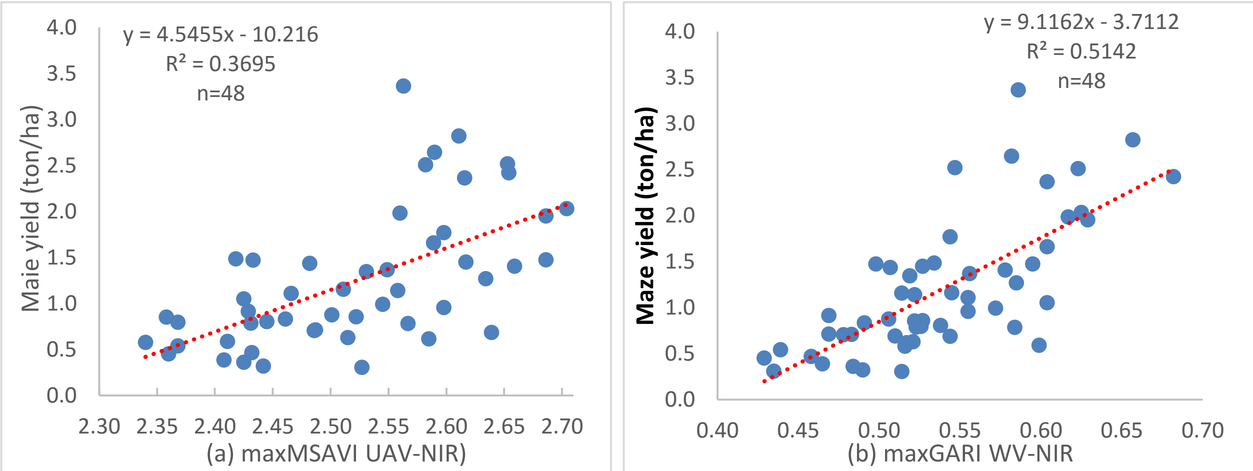


Figure 21: (a) Maize yield relationship with maxMSAVI derived from UAV-NIR during the maize growing season from sowing to senescence and (b) maxGARI derived from WV-NIR data

The results obtained from integrating WV and UAV indicate the coefficient of determination was 0.44, which was derived using single-date NDVI imagery at fruit development stage. This indicates a good potential for image integration in estimating maize yield at large scale by integrating very high spatial resolution imagery with coarse resolution data. However, advanced image fusion algorithm is recommended for enhancing the accuracy of the VI imagery integration, especially for multi-temporal imagery integration.

In summary, single-date WV vegetation index metrics performed better than UAV-RGB and NIR derived metrics. Likewise, the performance of single-date UAV-RGB camera VI metrics was better than UAV-NIR camera. However, cumulative UAV-NIR vegetation index data showed a better relationship than with single-date imagery during flowering and fruit development stage. There was general agreement between the three datasets on the best period for estimating maize yield to be during fruit development which occurs approximately 60-70 days from the sowing date.

Table 6: Summary of optimal VI indices and vegetation variables with corresponding R^2 and maize yield RMSE

Bands	Sensor	VI Metrix	Maize stage	Index	function	R^2	Adj.R^2	RMSE	n
RGB	UAV	Single image.	Fruit dev	VARI	Y= 11.609x+ 1.06	0.508***	0.489	0.483	28
NIR	UAV	Single image	Fruit dev	NDVI	Y=4.298x – 1.278	0.322***	0.307	0.600	50
NIR	UAV	Single image	Fruit dev	MSAVI	Y=5.394x – 5.417	0.322***	0.308	0.600	54
NIR	WV	Single image	Fruit dev	EVI	Y = 10.57x - 4.3502	0.613***	0.626	**0.460**	54
RGB	UAV	Variable	Silking-Fruit-dev.	VARI	Y=7.647x + 0.253	0.455***	0.434	0.532	28
NIR	UAV	Variable	Silking-Fruit-dev.	NDVI	Y=3.551x – 3.148	0.372***	0.359	0.577	50
NIR	WV	Variable	Silking-Fruit-dev.	GNDVI	Y=6.476x – 7.187	0.570***	0.562	**0.586**	54
RGB	UAV	MaxVARI	Entire season	VARI	Y=0.00022 + 1.049	0.031*	-0.006	0.750	28
NIR	UAV	MaxMSAVI	Entire season	MSAVI	Y=12.28x – 14.84	0.313***	0.299	0.603	50
NIR	WV	MaxGARI	Entire season	GARI	Y=9.116x – 3.711	0.514***	0.504	**0.500**	54

Level of statistical significance p***=0.01; p**=0.05; p*=0.1

n-number of sampled (varies depending on availability of NIR, RGB imagery)

4.2. Bootstrap model validation results

The bootstrap resampling validation computed using R-software was able to resample the data (n=54) 3600 times and generated normally distributed bootstrap sample using linear model derived from EVI index data and maize yield relation as shown in Figure 22. The histogram and the quantile plots indicate the sampled population distribution was normally sampled test.

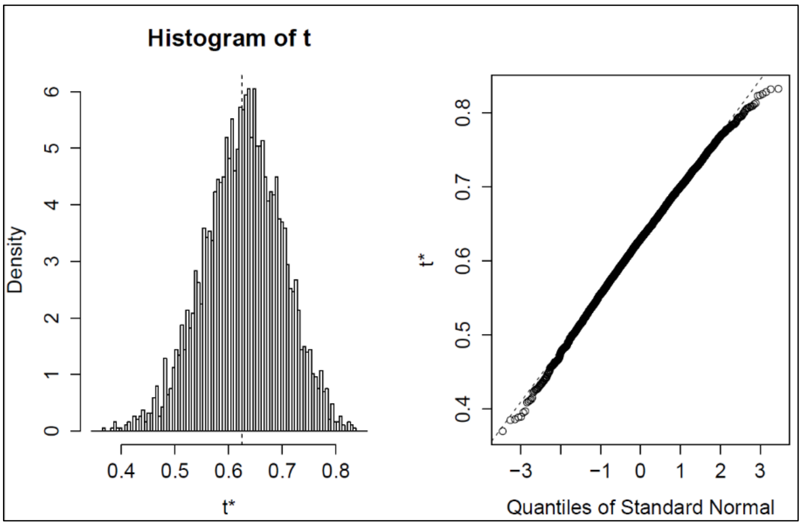


Figure 22: Normal distribution of the bootstrap sample population distribution shown in the histogram and the quartile plots computed from Enhanced Vegetation Index (EVI).

The results indicate clearly that EVI was the optimal index. The criteria used to select the optimal model was the model with high AdjR^2, RMSE, Standard Error (SE) and high lower and upper in bias-corrected and accelerated (BCa) at 95% confidence interval as summarized in Table 7. In addition, the model with a small sample (n=28) considering the samples were drawn from only one study site.

Table 7: Bootstrap result of maize yield-VI validation

VI metric	Equation	AdR^2	RMSE	SD	Bias	SE	BCa (95% CI	
							Lower bound	Upper bound
EVI	Y = 10.576x - 4.3502	0.601*	0.460	0.478	0.028	0.072	0.476	0.753
cumGNDVI	Y=6.476x – 7.187	0.562*	0.586	0.654	0.003	0.079	0.399	0.707
MaxGARI	Y=9.116x – 3.711	0.504*	0.500	0.093	0.006	0.092	0.290	0.664
Significant at $p < 0.00$ Results based on 3600 bootstrap samples								

4.3. Field-level maize yields variability

The yield variability map in Figure 23 shows spatial variability within and between fields. The variability map was computed using single-date WV-EVI equation during fruit development maize stage. The RMSE of the actual and predicted yield was 0.45 ton/ha with a bias error of zero. For the validation test carried out using Bootstrap (BCa, SDE, and SE) the model performed well. This indicates that the model is robust enough and could be applied for to an independent maize yield datasets. The equation that best described maize yield and VI relationship was Y = 10.576x - 4.35

Table 8 shows the model results in which at 95% confidence level, the actual and predicted maize yield has similar mean confirming the model good performance. However, the model had substantial effect on both the maximum and minimum maize yield whereby the model under predicted yield. This can be further confirmed by the negative intercept (-4.35). These results are important when interpreting the computed maize yield maps.

Table 8: Descriptive statistics of the actual and predicted maize yield (ton/ha)

	Actual yield	Predicted yield
Mean	1.201	1.201
Variance	0.526	0.333
Std. Deviation	0.725	0.576
Minimum	0.310	0.190
Maximum	3.370	2.590
95% confidence of the mean		

The yield variability map (Figure 23) shows high spatial variability within and between the fields. The highest maize yield (green) in most cases were fields with maize mixed with pigeon peas or sunflower. The green patches in the fields do not represent yield, rather they are isolated trees within the fields. The reason for such inaccuracies is the effect on non-maize vegetation growing in the maize yield which increases VI. The other factor that contributed to difference in maize yield variability is that some of the maize in some fields were already in senescence stage, especially those farmers who planted in late January or early February.

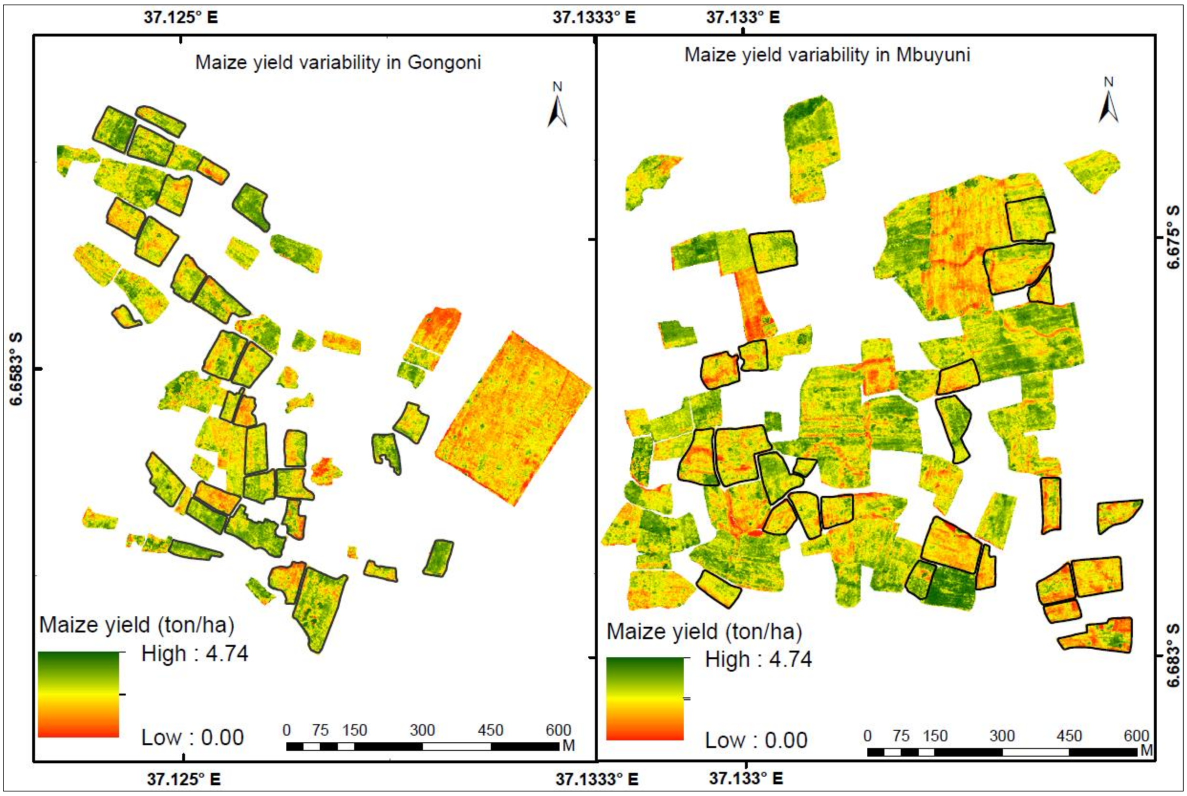


Figure 23: Pixel based result from modeling maize yield variability using Enhanced Vegetation index (EVI) derived from WorldView-2 imagery acquired during flowering maize stage. The sampled maize fields are indicated with black boundaries.

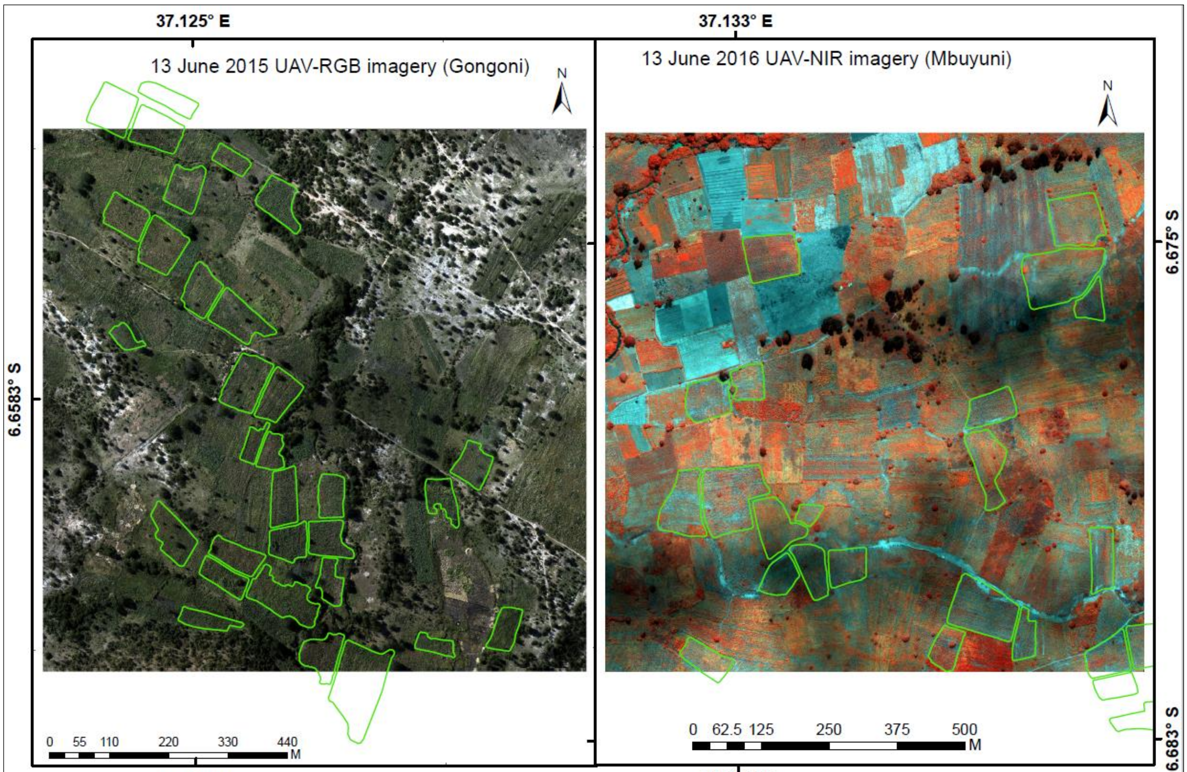


Figure 24: Fine spatial resolution UAV-RGB and UAV-NIR imagery (0.05m) acquired on 13 June 2015 showing maize field during flowering stage in the two study sites (bright green polygons are the fields sampled)

The graph shows that the model over-predicted maize yield especially the low which can be attributed to non-maize vegetation growing in the maize yield and which had high reflectance values. This can be observed with the maize yield scatter graph in which most of the predicted maize yield lie above the scatter plot line. This is further supported by the descriptive statistics indicated previously in
Table 8 where the predicted and actual yield had same mean but the different standard deviation in which predicted maize yield showed the least variation as compared to the actual maize yield. The scatter plot in Figure 25 indicates the spread of the predicted maize yield as compared to the actual maize yield obtained during the interview with the farmers.

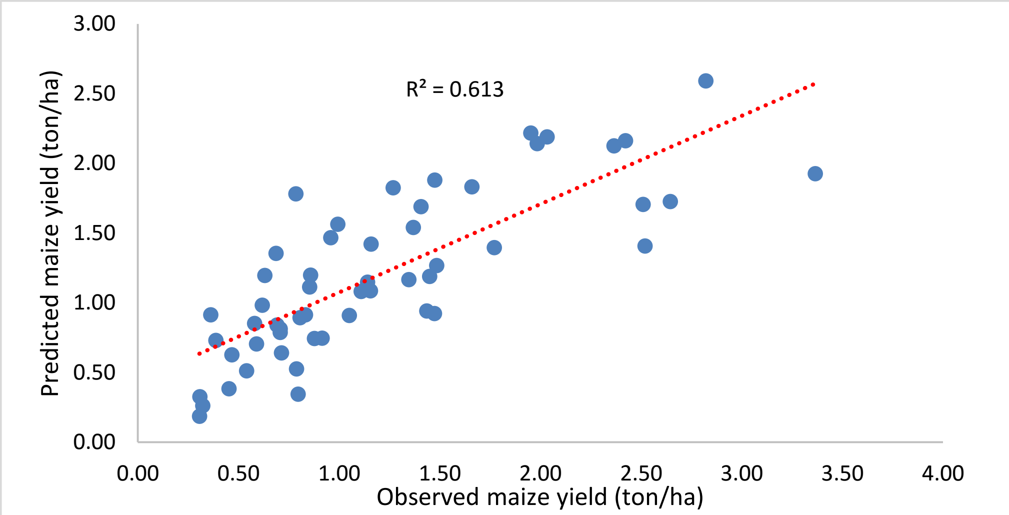


Figure 25: Comparison of reported yield with values predicted from WV-EVI at fruit development stage Effect of management factors on maize yield

4.4. Effect of management factors on maize yield

A further test was carried out to determine if differences in management factors reported by the farmers had a significant influence on yield. In this regard, Analysis of Variance (ANOVA) was computed based on different management factors which include the method of tilling applied, cropping system, seeds planted, the number of times weeding was carried out, difference in planting dates and level of pest and diseases infestation on maize crops. The ANOVA results summarized in Table 9 shows the only observable difference was the number of times weeding was carried out and the method of tilling applied. However the ANOVA test does not show which factor had significant impact on yield. The hypothesis was if ANOVA the computed p-value is less than alpha (α=0.05) then the management factor would be statistically significant at 95% (0.05) confidence level (i.e. $p< \alpha$)

Table 9: ANOVA results of interview field management practices on maize yield

Management factor	Interview results	Grouping	Conclusion
Tilling methods	0.03< 0.05	Tractor, oxen & hand hoe	Different
Cropping system	0.34>0.05	Mono-cropping, mixed cropping &intercropping	No Difference
Seeds planted	0.59>0.05	Local & certified seeds	
Weeding	0.04> 0.05	Once or twice	Different
Planting date	0.16>0.05	Grouped in 10 days difference	No difference
Pest	0.18 >0.05	High, Average and Low	No difference
Diseases	0.34 > 0.05	High, average and low	No difference

Separately, Fisher's Least Significant Difference (LSD) test was carried out to determine which of the tilling methods had a significant effect on yield. The results indicate the use of tractor and hand hoe had a significant effect on yield ($0.02<\alpha >0.05$), although the test does not indicate which gives a higher or low yield. However averaging the yield of farmers who reported having used tractors as compared to those who used handhoe, those who used tractor harvested more yield than those who used hand hoes or oxen.

Although it has been established in many studies that planting dates have a significant influence on yield, the probable case in this as to why there was no significant difference in yield could be due to the fact that most farmers planted around the same time which could not make much difference in maize yield. In addition, other factors may have contributed to yield difference more than effects of planting. The few who planted early or late may also have had other factors which influence maize yield much more than the planting dates. In regard to weeding, the effect of weeding of competing for nutrients with maize plant makes the maize crop weak. In addition, weeds attract pests and diseases which attack maize crops resulting in low yield. However, it is interesting to note there was no much difference between the farming systems against the expectation that mono-cropped fields would have a higher yield as compared to mixed or intercropping. The probable reason could be in the mono-cropped fields other factors that affect yield played significant role thereby reducing the maize yield. Pest and diseases effect as reported by many farmers was not a major problem as compared to weeds and this is the reason why the effect of pests and diseases had no significant effect on yield.

A further test to determine spatial field-level maize yield variability was tested using the high average-low yield location reported to test if it was corresponding to the results predicted by the VI-yield model. The results of non-parametric Spearman's correlation gave a result of $R^2=0.202$ (n=920). Although this is extremely low compared to the yield estimate, the information which can be deduced from these result is that for most of the points reported by the farmers does not correlate with the VI obtained. The reason is further justified by the box plot in Figure 22 which indicates the variation of the reported maize yield rank in order of high medium and low yield by the farmer interviewed against the estimated maize yield. The section of the fields reported having low yield seems to have the highest variation. This can be attributed to the VI images predicting high yield in areas with trees and weeds while the actual yield is low. The box plot showing high yield had the lowest variation meaning whatever location the farmer reported having harvested high yield corresponded relatively well with what was observed in the imagery. Lastly, the average yield variation was slightly higher than high yield variation which is due to some section reported having low yield when compared to the predicted maize yield as shown with the high reported yield having the least variation.

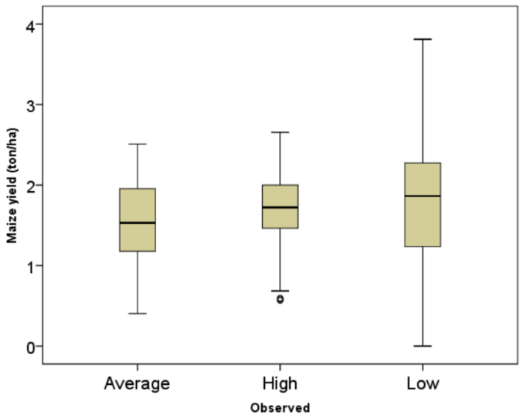


Figure 26: Box plot showing differences in-field reported maize yield level in comparison to the yield at derived yield variability map.

5. DISCUSSION

5.1. Assesment of yield using fine spatial resolution data

Fine spatial resolution imagery captures fine structures of maize plant such as the leaves while coarse resolution data cover canopy level. In their assessment of effect of spatial resolution on maize yield relationship, Geipel et al., (2014) found that by varying spatial resolution from fine (0.02 m) to intermediate (0.04 m) and fine resolution (0.06 m); the inter-intermediate and fine resolution ExG index relationship with maize yield had a better R^2 than very fine spatial resolution data in which they attributed to high noise from soil and non-maize vegetation. However, for coarse resolution data, the R^2 is degraded by the mismatch between maize fields and the pixel sizes especially for the 250 meter resolution data when used in highly fragmented fields (Duncan et al., 2015). Therefore, in determining this VI-maize yield relationship, it is important to consider the stage of maize growth and the type of vegetation cover existing within the maize yield.

In addition, fine spatial resolution RGB imagery is very useful for visualization and for accurate delineation of maize harvested area as shown in the study. Fine spatial resolution helps in mitigating the challenge of coarse resolution data which does provide sufficient resolution for delineating maize fields (Lobell, 2013). However, automatically delineating boundaries would still be a challenge considering fuzzy maize yield boundaries which are further complicated by farmers who change their land use in the middle of the season due to such factors as poor performance of the maize. Furthermore, the effect of non-crop vegetation in the crop field has been shown to have significant impact on coarse spatial resolution data and is still a challenge for the fine spatial resolution imagery (Chen et al., 2008). This had a significant effect on the maize yield relationship derived in this study. To mitigate this, there is a need to separate crop and non-crop vegetation using such methods as of spectral thresholding as applied by Ridler et al., (1978) than conventional classification approaches which is limited by the need of validation points (Rembold et al., 2013). In their review of the use of remote sensing in yield gap analysis, Duncan et al., (2015) note the challenge of effect of non-crop vegetation not only in regard with coarse resolution data, but also with fine resolution imagery though the effect is not comparable to the coarse resolution data. Furthermore, there using crop texture especially with multi-temporal images will provided a better method of detecting weeds and other non-maize crops. In addition, other methods such as combining VI data crop height model which is obtained by subtracting digital surface model (DSM) and digital elevation model (DEM) (Geipel et al., 2014).Therefore, for accurate yield assessment fine-spatial resolution remote sensing data still holds the key to achieving improve maize yield assessment at field level, only if non-maize crop are masked out.

The limitation with very-fine-resolution data is the temporal frequency of acquisition and allows monitoring only at small scale. However, efforts have been made to fuse high resolution and coarse resolution so as to generate high-resolution synthetic VI image which would allow monitoring of the large area (Boschetti et al., 2015). The simple linear regression tested integration technique applied in this study show there is the potential for integrating both WV and UAV images. However this calls for used of improved algorithms such as Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM) Gevaert et al., (2014) or such algorithm described by (Zurita-Milla et al., 2011). The advantage with these methods is that it takes into consideration high temporal and spatial variability in vegetation reflectance.

Accurate pre-processing of fine spatial resolution data is very important for accurate maize yield estimation from fine spatial resolution data. Horizontal alignment errors in fine resolution imagery have a large impact on the displacement of features considering its fine spatial resolution. In order to account for horizontal misalignment, Geipel et al., (2014) suggest the use of polygon containing field information for realigning all the images acquired during the crop season. Combining digital terrain model (DTM) with an accuracy of 10-15 cm and ground control points (GCP) of the approximate accuracy of ~30 cm acquired automatically during UAV flight has shown to substantially improve the accuracy of geometric and orthorectification of UAV imagery (Vallet et al., 2011). Furthermore, Geipel et al., (2014) suggest the use of DEM and DTM than the use of dense point cloud which can be hardware demanding task. In regard to WV, the geometric errors are large due to large area coverage by the sensor than UAV and considerable care has to be considered when comparing results of these two datasets (or even integrating). This was noted in this study since there was uniform shift between WV and UAV which was corrected by shifting pixel position. The limitation with this approach is that it does not give the RMSE of the location shift which can be used to evaluate the accuracy of geometric correction. Alternative to this would be to use GCP acquired from the UAV which was not readily available for use during this study. In regard to WV imagery, the effect of atmospheric effect have an impact on the signal detected by the sensors. However, in as much as the effort is made in correcting atmospheric and radiometric bias, errors due to the row-based cultivation of corn and missing canopy, early stages require very high spatial resolution which gets less important as the maize canopy cover increase (Geipel et al., 2014). Accurate estimation of maize yield at field level is dependent among other variable spatial resolution, vegetation index and field management factors. The study showed that while the use of fine spatial resolution vegetation metrics has the potential to improve maize yield estimation, it is complicated by non-uniform field management practices.

5.2. Statistical emperical model use in yield assessment

There are several empirical models that have been applied in modeling VI-yield relationship. Of these models, linear models have shown to be an optimal model. However, the performance of the model depends on the quality of yield data used given there is the possibility of non-negligible errors in farmer reported data which have an impact on model performance (Lobell, 2013). These errors might have contributed to the model showing low EVI for high yield which is contrary to what the model was presenting. The distribution of the sample should also be considered as this has an impact on model performance. Although a number of techniques are available for validating empirical yield models, the conventional splitting of the data into two sets (test and training) does not give a good indication of model performance considering large sample is required for it to be split. In addition, the data has to be normally distributed otherwise the effect of outliers would have a significant impact on model results. In order to overcome this cross-validation and bootstrap resampling techniques are preferred in computing residual errors in the model. The advantage of cross-validation is that validation data is different from the training data, however, it is limited in the fact that the sample has to be substantially large for it to be divided into training and testing sample and furthermore, the model has high variability which changes when a new sample is drawn. On the other hand, bootstrap does not require data transformation in case the data is not normally distributed, however, the pick one with replacement has been found that almost 30% of the samples drawn also form part of the model (Koehler et al., 2009). In general, all regression models perform well, however, the linear model is preferred in yield-VI modeling. In addition, normalizing data using statistical transformation changes the original yield values and therefore resampling techniques such as bootstrapping and cross-validation is preferred.

5.3. Fine spatial resolution vegetation metrics use in crop yield assessment

The remote sensing approach of using VI metrics is based on the fact that vegetation reflectance provides a measure of amount and condition of greenness which in turn is applied which is a proxy used in estimating yield (Duncan et al., 2015). In this regard, use of very-fine resolution detects any green vegetation in the field which may lead to erroneous interpretation of yield variability maps. A good example of this study was with sunflower and pigeon peas planted fields had significantly high VI even after maize senescence considering they were planted 2-4 weeks after maize. Using the current modeling approach, the VI of sunflower or weeds will be considered and this leads to overestimation of yield.

The vegetation indices computed from RGB camera carries limited spectral information as compared to NIR camera due to the high atmospheric effect on the visible band (Lebourgeois et al., 2008). In the study, EVI showed a high predictive power in maize yield estimation as compared to other indices tested. In other studies reviewed, EVI has shown high predictive in maize yield estimation (M. Wang et al., 2014; Zhang et al., 2014). Considering EVI uses 3 bands as compared to NDVI, this has been shown to have shown better results since the blue band is known to provide atmospheric correction as compared to NDVI (Bolton et al., 2013). EVI, in this case, had a higher R^2 (0.63) as compare to a study by M. Wang et al., (2014) which they obtained $R^2 = 0.43$, and lower than Bolton et al., (2013) in which they obtained $R^2=0.67$. The difference between this study and the studies highlighted is the geographical zones in which the study was carried out where in the case of largest EVI was carried out in the United States of America while the lowest EVI results were carried out in India. Furthermore, this study used field be attributed to the fact in this study, field aggregated VI was used with fine spatial resolution imagery as compared to the two which used coarse resolution data and classified crop map.

Maize yield is an end product but maize crop undergoes through a number of stages to produce yield. Therefore, understanding when yield components can be determined is important in interpreting management and environmental factors that influence maize yield (Darby et al., 2013). Determining the optimal stages of maize growth upon which maize yield can be estimated using remotely sensed vegetation indices metrics was one of the key focus of the study. As indicated in the results, an optimal period with high maize yield relationship was found between 60-70 days of maize development. This corresponds to silking and fruit development which is in agreement with a number of maize yield studies (Omoyo et al., 2015; M. Wang et al., 2014). Although these studies was carried out in different geographic set up with different datasets, there seemingly to be agreement on the optimal period to be during silking stage.

The regression-based model developed in this study was empirically derived using field-level interview data and a test carried out on various very fine spatial resolution vegetation metrics. The use of single-date imagery shows to be promising for maize yield estimation at field level. In order to accurately predict maize yield using single date imagery, the timing of the maize stage is very important for optimal maize yield estimation. There is a general consensus among researcher that the optimal period of predicting maize is from flowering to fruit development which 60-70 days) (Bolton et al., 2013). The result of this study, is in agreement with the period of maize yield estimation as established in literature cited. The challenge with single-date images is the difficulty of getting a cloud-free imagery, especially in areas where cloud cover is a problem. An alternative would be to use UAV around fruit development stage as this shows improved relationship with yield. Since UAV images are affected by shadows due to its high resolution, the timing should be before mid-day and preferably using GLI index if RGB camera is to be used and GRVI with NIR camera as these two indices seem to explain more than 47% of the maize yield variation. In terms of high prediction power, WV imagery using EVI seems to be an optimal option.

Although the results of CumVI is almost same as for the single-date around the stage of anthesis and fruit development, CumVI seems to give a lower RMSE as compared to single-date or MaxVI. This indicates two observation, first, the cumulating of VI over the period between anthesis and fruit development captures the events that occur during the critical stages of grain formation in maize plant(Viña et al., 2004). Secondly, the changes in cumVI is a result of factors such as pests, diseases, and extreme weather conditions will bring about changes in VI which make cumulative index give a low RMSE. Vegetation index (VI) accumulation at the beginning of the maize growing season showed weak relationship for both the UAV and WV data and this suggests that for the optimal prediction using cumulative model, the inclusion of early season affect model accuracy as compared to late season (Bakhsh et al., 2000; Basso et al., 2013). The use of UAV RGB and NIR shows little difference although NIR is preferred as it gives a low RMSE in estimating maize yield. The performance of MaxNDVI was poor especially with UAV RGB and NIR imagery. This can be attributed to the effect of weeds which seems to be less detected with RGB camera than NIR. The WV derived maximum indices performed better than UAV given most of the indices could explain 40% of the variation.

The optimal spectral index based on this study is EVI which is similar to a maize yield study by Bolton et al., (2013) in which EVI outperformed NDVI (R^2=0.58 against NDVI (R^2=0.53). Its performance was constantly high in both UAV and WV images. This indicates that different VI has different strengths in predicting maize yield. The difference between the two indices (EVI and NDVI) is that EVI which is more sensitive to canopy structure and variables such as leaf area index while NDVI is more sensitive to chlorophyll content in plant leaves (Huete et al., 2002). Furthermore, it has been found in a number of studies that NDVI saturates with dense canopy cover and maintains this high values throughout the cropping season as compared to EVI (Wardlow et al., 2007) During maize development, the unfavourable conditions in the grain filling period (anthesis and physical maturity) has been found to likely impair pollination and reduce the fertilized kernels that are destined to be filled(Viña et al., 2004). Maize phenology is divided majorly into vegetative (emergence to tasselling according to a number of leaves) and reproductive (silking to physiological maturity according to the degree of kernel development). Within these stages, several transition is important in terms of management. During maize development, the maximum yield can be realized if there is sufficient supply of nutrients under favorable condition (i.e. soil moisture, solar radiation, and temperature).Unfavourable conditions at the beginning of the reproductive cycle (tasselling and anthesis) are likely to impair pollination and reduce the number of fertilized kernels that are to be filled (Viña et al., 2004). Any adverse condition during the grain filling period (between anthesis and fruit development) are likely to impair pollination. Detecting early onset of senescence is important because it can have a direct influence on yield. The flowering and grain filling periods are the most critical for most crops; any water stress during these crop growth stages may result in reduced grain yields (Mkhabela et al., 2011).

5.4. Field level maize yields variability

The variation in date of planting is important in that maize planted early will be in different stages of development as compared to those planted late. However, the spectral information captured by a single image will be measuring spectral information from different stages of development and hence can influence the model accuracy. Optimum maize production calls for the good timing of the planting dates. Postponing planting dates has been found to have significant negative effect on maize yield (Azadbakht et al., 2012)
It is important to note the link between yield estimation and biomass. Weeds control has a significant effect on weed density whereby if there is no weed control the density of weed tend to be high (Udom et al., 2010). Weed management options have shown to have significant effects on weed suppression, maize height and dry grain yield of maize (Joshua et al., 2008).This is because weeds indirectly affects maize cob length, cob diameter, and number of grain per cob and dry grain yield in fields with a lot of weeds is attributed serious competition of weeds with maize plants for soil water nutrients resulting in reduced plant height and maize yield.

Although the results from farmer's interview indicate farming systems had no significant relationship with maize yield, visual interpretation of the VI images and the resulting maize yield map indicates fields with pigeon peas and sunflower had consistently high VI while comparing to the reported yield there was an indication of over prediction in the images. The second evidence is the results obtained from yield ranking in which the R^2=0.202 (p>0.01) which indicate most of the e evidenced by the cropping systems has a significant impact on maize yield estimation. Although there was no direct test to determine the effect of soil, visually comparing the soil map in Figure 4, with the maize yield variability map, there is some indication that soil type may have contributed to yield difference considering that most farmers in both sites did not use fertilizer. In summary, the fine-resolution imagery has shown areas of targeted intervention. This is important for better management practices especially for areas where the yield level was low

5.5. Effect of management factors on maize yield

Maize growth stage has an effect on VI-maize yield relationship. However determining exact stage of maize development is difficult considering the difference in planting dates, management factors such as weeding, maize varieties. In this regard, the use of cumulative vegetation index would come in handy in reducing the difference in maize growing stage. Best time for predicting maize yield using multi-temporal VI data has been established to be between 50-70 days after planting date (M. Wang et al., 2014). Although the strongest correlation between yield and NDVI has been found around maximum VI in a study by Tucker (1980), Maximum VI in this case and a number of other studies has shown weak correlation with maize yield (M. Wang et al., 2014). In other studies, MaxVI has been shown to have varying peak correlation with yield during the season (M. Wang et al., 2014) and this could explain why the maxVI was inconsistent between the UAV and WV data given the imagery used were acquired in different periods. It has been established that yield-VI relationship varies as a function of time during the growing season.

6. CONCLUSION AND RECOMMENDATION

The study demonstrated the potential of using fine resolution data in assessing maize yield at field level. In this thesis, several vegetation indices metrics were tested to determine the optimal vegetating index which was found to be EVI derived from WorldView imagery at the silking stage. Furthermore, the study found that cumGNDVI outperformed maxGARI in estimating maize yield indicating that there are of factors that affect maximum vegetation greenness relationship with maize yield. One of the observations made in the study was the effect by non-maize vegetation grown in the maize field. The period before maximum greenness showed the least maize yield relationship which was attributed to minimal vegetation cover. During fruit development period, the VI-yield declined as the maize headed toward senescence which was attributed to decrease in green biomass a result of a decrease in photosynthetic activity in the maize crop. The study found out that WV derived indices performed better than UAV indices. The plausible explanation was the difference in the image acquisition dates and (although not directly tested) differences in spectral bandwidth in which WV had a narrow bandwidth as compared to UAV. An effort was made to integrate same date UAV derived NDVI and WV NDVI during the flowering period. The result indicated a good potential for integrating airborne UAV derived imagery with satellite-based WV images for local or regional scale assessment of maize yield. However, use of an advanced algorithm which gathers for temporal variation in VI is recommended in the case of different date's image integration. The use of bootstrap resampling technique applied in model validation resulted in the selection of an optimal model that was used to derive yield variability map which further proves its ability to provide good statistical validation measures. The yield variability map showed high yield variation between low yield and high yield fields. However, it was noted that the variability was contributed by the difference in the dates the image was acquired (considering single data image was used) and secondly, differences in stages of maize growth. Thirdly, it was the confounding effect of non-maize vegetation growing in the maize field which overestimated yield. A confirmation of the effect of non-maize vegetation on maize yield variability map was noted when Spearman's rank correlation test was applied in correlating field-level collected data and the actual output yield which resulted to the very weak relationship (r^2=0.2).In terms of the effect of management factors, the number of times a maize field was weeded and method of tilling applied showed a significant relationship with yield. Other management factors such as planting date, crop pests and diseases, cropping systems and source of seeds planted showed no significant effect on yield.

However, the results obtained in the study is not all that good considering the model could not explain all the maize yield distribution adequately. This was shown by the scatter plot in Figure 20 whereby high yield corresponded to low EVI which was not the actual case of what the overall model was depicting. The reason for such occurrence was the uncertainty in the quality of field collected production data and also the difference in planting dates which contributed to differences in average spectral VI within the fields. Although fine spatial resolution in yield estimation provides great potential for estimating maize yield at field level, study could not establish clear difference based on the result obtained. This was largely affected by heterogeneous vegetation cover in the field which affected yield estimation considering green biomass was used as a proxy for yield estimation which does not have direct link with yield.

In order to improve crop yield assessment using fine spatial resolution imagery in the future it is recommended that: - First, accurate maize production and delineated harvested area is used, preferably destructive sampling rather than the interview data (2). There is need to accurately classify maize and non-maize pixel using such methods as VI thresholding, use of texture and combining crop height model. Thirdly, accurate alignment between UAV and WV imagery should be carried out to avoid pixel location shift and fourthly, due consideration of planting dates as it reflectance values changes as crop grows.

REFERENCES

Arellano, A. A. P. (2015). Sensores. *Sistemas Y Soluciones de Mapeo Aéreo Profesional (SYSMAP).*

Atzberger, C. (2013). Advances in remote sensing of agriculture: Context description, existing operational monitoring systems and major information needs. *Remote Sensing*, *5*(2), 949–981. http://doi.org/10.3390/rs5020949

Bakhsh, A., Jaynes, D. B., Colvin, T. S., & Kanwar, R. S. (2000). Spatio-temporal analysis of yield variability for a corn-soybean field in Iowa. *Transactions of the American Society of Agricultural Engineers*, *43*(1), 31–38. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-0034104666&partnerID=tZOtx3y1

Basso, B., Cammarano, D., & Carfagna, E. (2013). Review of Crop Yield Forecasting Methods and Early Warning Systems. *The First Meeting of the Scientific Advisory Committee of the Global Strategy to Improve Agricultural and Rural Statistics*, 1–56.

Bolton, D. K., & Friedl, M. A. (2013). Forecasting crop yield using Remotely Sensed vegetation indices and crop phenology metrics. *Agricultural and Forest Meteorology*, *173*, 74–84. http://doi.org/10.1016/j.agrformet.2013.01.007

Boschetti, L., Roy, D. P., Justice, C. O., & Humber, M. L. (2015). MODIS–Landsat fusion for large area 30m burned area mapping. *Remote Sensing of Environment*, *161*, 27–42. http://doi.org/10.1016/j.rse.2015.01.022

Brown, M. E., & de Beurs, K. M. (2008). Evaluation of multi-sensor semi-arid crop season parameters based on NDVI and rainfall. *Remote Sensing of Environment*, *112*(5), 2261–2271. http://doi.org/10.1016/j.rse.2007.10.008

Brown, M. E., de Beurs, K., & Vrieling, A. (2010). The response of African land surface phenology to large scale climate oscillations. *Remote Sensing of Environment*. http://doi.org/10.1016/j.rse.2010.05.005

Chen, Z., Li, S., Ren, J., Gong, P., Zhang, M., Wang, L., … Jiang, D. (2008). Monitoring and management of agriculture with Remote Sensing. In S. Liang (Ed.), *Advances in Land Remote Sensing System: Modeling, Inversion and Application* (pp. 401–425). University of Maryland, USA: Springer.

Claverie, M., Demarez, V., Duchemin, B., Hagolle, O., Ducrot, D., Marais-Sicre, C., … Dedieu, G. (2012). Maize and sunflower biomass estimation in southwest France using high spatial and temporal resolution remote sensing data. *Remote Sensing of Environment*, *124*, 844–857. http://doi.org/10.1016/j.rse.2012.04.005

Darby, H., & Lauer, J. (2013). Critical Stages in the Life of a Corn Plant. *Plant Physiology*, 15–20. Retrieved from http://corn.agronomy.wisc.edu/Management/pdfs/CriticalStages.pdf

de Beurs, K. M., & Henebry, G. M. (2005). A statistical framework for the analysis of long image time series. *International Journal of Remote Sensing*, *26*(8), 1551–1573. http://doi.org/10.1080/01431160512331326657

de By, R. A., & Zurita-Milla, R. (2015). *STARS : Rich remote sensing data for tropical smallholder farming: STARS : Rich remote sensing data for tropical smallholder farming : powerpoint. Presented at: SIKS / Twente data science colloquium, 20 april 2015, Enschede, Netherlands. 52 slides*. Retrieved from http://www.stars-project.org/.system/dl/ec~AQJGABIAVRoFAgY6CwCJRPIBOqOI/STARS_Fact_sheet.pdf?whs-download=STARS_Fact_sheet.pdf

Duncan, J. M. A., Dash, J., & Atkinson, P. M. (2015). The potential of satellite-observed crop phenology to enhance yield gap assessments in smallholder landscapes. *Frontiers in Environmental Science*, *3*(August), 1–16. http://doi.org/10.3389/fenvs.2015.00056

Efron, B., & Gong, G. (1983). A leisurely look at the bootstrap, the jackknife, and cross-validation. *American Statistician*, *37*(1), 36–48. http://doi.org/10.1080/00031305.1983.10483087

Erenstein, O., Kassie, G. T., Langyintuo, A., & Mwangi, W. (2011). *Characterization of Maize Producing Households in Drought Prone Regions of Eastern*. Mexico.

European Space Agency. (2015). Sentinel-2 - Missions - Sentinel Online. Retrieved August 12, 2015, from https://sentinel.esa.int/web/sentinel/missions/sentinel-2

Everaers, P. (2010). The present state of agricultural statistics in developed countries: situation and challenges. In R. Benedetti, M. Bee, G. Espa, & F. Piersimoni (Eds.), *Agricultural Survey Methods* (1st ed., pp. 1–24). United Kingdom: John Wiley & Sons Ltd.

FAO, IFAD, & WFP. (2014). *The State of Food Insecurity in the World 2014. Strengthening the enabling environment for food security and nutrition.* Rome, Italy: FAO.

Funk, C., & Budde, M. E. (2009). Phenologically-tuned MODIS NDVI-based production anomaly estimates for Zimbabwe. *Remote Sensing of Environment*, *113*(1), 115–125. http://doi.org/10.1016/j.rse.2008.08.015

Geipel, J., Link, J., & Claupein, W. (2014). Combined Spectral and Spatial Modeling of Corn Yield Based on Aerial Images and Crop Surface Models Acquired with an Unmanned Aircraft System. *Remote Sensing*, *6*(11), 10335–10355. http://doi.org/10.3390/rs61110335

Gevaert, C. M., Tang, J., Suomalainen, J., & Kooistra, L. (2014). Combining Hyperspectral UAV and Multispectral FORMOSAT-2 Imagery for Precision Agriculture Applications, 2–5.

Gitelson, A. A., Kaufman, Y. J., Stark, R., & Rundquist, D. (2002). Novel algorithms for remote estimation of vegetation fraction. *Remote Sensing of Environment*, *80*(1), 76–87. http://doi.org/10.1016/S0034-4257(01)00289-9

Groote, H. De, Doss, C., Lyimo, S. D., & Mwangi, W. (2002). Adoption of maize technologies in East Africa – What happened to Africa's emerging maize revolution? A paper presented at the FASID Forum V. In *"Green Revolution in Asia and its transferability to Africa,."* Tokyo.

Haboudane, D., Miller, J. R., Tremblay, N., Zarco-Tejada, P. J., & Dextraze, L. (2002). Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote Sensing of Environment*, *81*(2-3), 416–426. http://doi.org/10.1016/S0034-4257(02)00018-4

HarvestChoice. (2010). Rainfall Variability and Crop Yield Potential | International Food Policy Research Institute, Washington, DC., and University of Minnesota, St. Paul, MN. Retrieved July 29, 2015, from http://harvestchoice.org/labs/rainfall-variability-and-crop-yield-potential

Henrich, V., Krauss, G., Götze, C., & Sandow, C. (2012). Development of an online indices database: Motivation, concept and implementation. Retrieved August 10, 2015, from http://www.indexdatabase.de/db/is.php?sensor_id=40

Hoefsloot, P., Ines, A., Dam, J. Van, Duveiller, G., Kayitakire, F., & Hansen, J. (2012). *Combining crop models and remote sensing for yield prediction: Concepts, applications and challenges for heterogeneous, smallholder environments*. *JRC Scientific and policy reports*.

Honkavaara, E., Saari, H., Kaivosoja, J., Pölönen, I., Hakala, T., Litkey, P., … Pesonen, L. (2013). Processing and Assessment of Spectrometric, Stereoscopic Imagery Collected Using a Lightweight UAV Spectral Camera for Precision Agriculture. *Remote Sensing*, *5*(10), 5006–5039. http://doi.org/10.3390/rs5105006

Huang, J., Wang, X., Li, X., Tian, H., & Pan, Z. (2013). Remotely Sensed Rice Yield Prediction Using Multi-Temporal NDVI Data Derived from NOAA's-AVHRR. *PLoS ONE*, *8*(8), 1–13. http://doi.org/10.1371/journal.pone.0070816

Huete, a., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, *83*(1-2), 195–213. http://doi.org/10.1016/S0034-4257(02)00096-2

ISRIC - World Soil Information. (2015). Soil property maps of Africa at 250 m | ISRIC World Soil Information. Retrieved November 10, 2015, from http://www.isric.org/data/afsoilgrids250m

Jain, M., Mondal, P., DeFries, R. S., Small, C., & Galford, G. L. (2013). Mapping cropping intensity of smallholder farms: A comparison of methods using multiple sensors. *Remote Sensing of Environment*, *134*, 210–223. http://doi.org/10.1016/j.rse.2013.02.029

Johnson, B., Tateshi, R., Kobayashi, T., Tateishi, R., & Kobayashi, T. (2012). Remote sensing of fractional green vegetation cover using spatially-interpolated endmembers. *Remote Sensing*, *4*(9), 2619–2634. http://doi.org/10.3390/rs4092619

Jordan C.F. (1969). Derivation of leaf area index from quality of light on the forest floor. *Ecology*, *50*, 663–666. http://doi.org/doi: http://dx.doi.org/10.2307/1936256

Kajembe, G. C., Silayo, D. S. a., Mwakalobo, A. B. S., & Mutabazi, K. (2013). The Kilosa District REDD+ pilot project, Tanzania A socioeconomic baseline survey.

Khan, M. R., de Bie, C. a J. M., van Keulen, H., Smaling, E. M. a, & Real, R. (2010). Disaggregating and mapping crop statistics using hypertemporal remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, *12*(1), 36–46. http://doi.org/10.1016/j.jag.2009.09.010

Koehler, E., Brown, E., & Haneuse, S. J.-P. A. (2009). On the Assessment of Monte Carlo Error in Simulation-Based Statistical Analyses, *63*(2), 155–162. http://doi.org/10.1198/tast.2009.0030.On

Lebourgeois, V., Bégué, A., Labbé, S., Mallavan, B., Prévot, L., & Roux, B. (2008). Can Commercial Digital Cameras Be Used as Multispectral Sensors? A Crop Monitoring Test. *Sensors*, *8*(11), 7300–7322. http://doi.org/10.3390/s8117300

Lewis, J. E., Rowland, J., & Nadeau, a. (1998). Estimating maize production in Kenya using NDVI: Some statistical considerations. *International Journal of Remote Sensing*, *19*(13), 2609–2617. http://doi.org/10.1080/014311698214677

Lin, A. Y. M., Novo, A., Har-Noy, S., Ricklin, N. D., & Stamatiou, K. (2011). Combining GeoEye-1 satellite remote sensing, UAV aerial imaging, and geophysical surveys in anomaly detection applied to archaeology. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *4*(4), 870–876. http://doi.org/10.1109/JSTARS.2011.2143696

Livingston, G., Schonberger, S., & Delaney, S. (2011). Sub-Saharan Africa : The state of smallholders in agriculture. In *Conceference on New Directions for Smallholder Agriculture 24-25 January, Rome IFAD HQ* (pp. 1–31).

Lobell, D. B. (2013). The use of satellite data for crop yield gap analysis. *Field Crops Research*, *143*, 56–64. http://doi.org/10.1016/j.fcr.2012.08.008

Magehema, A. O., Chang, L. B., & Mkoma, S. L. (2014). Implication of rainfall variability on maize production in Morogoro , *4*(5). http://doi.org/10.6088/ijes.2014040404547

Marques da Silva, J. R., & Silva, L. L. (2008). Evaluation of the relationship between maize yield spatial and temporal variability and different topographic attributes. *Biosystems Engineering*, *101*(2), 183–190. http://doi.org/10.1016/j.biosystemseng.2008.07.003

Meier, U. (2001). Growth stages of mono-and dicotyledonous plants. BBCH Monographie, 166.

Meroni, M., Marinho, E., Sghaier, N., Verstrate, M., & Leo, O. (2013). Remote Sensing Based Yield Estimation in a Stochastic Framework — Case Study of Durum Wheat in Tunisia. *Remote Sensing*, *5*(2), 539–557. http://doi.org/10.3390/rs5020539

Mkhabela, M. S., Bullock, P., Raj, S., Wang, S., & Yang, Y. (2011). Crop yield forecasting on the Canadian prairies using MODIS NDVI data. *Agricultural and Forest Meteorology*, *151*(3), 385–393. http://doi.org/10.1016/j.agrformet.2010.11.012

Nathan, M. (2014). Food production variability and modeling in East Africa. Retrieved June 1, 2015, from https://www.agriskmanagementforum.org/content/food-production-variability-and-modeling-east-africa

Omoyo, N. N., Wakhungu, J., & Oteng'i, S. (2015). Effects of climate variability on maize yield in the arid and semi arid lands of lower eastern Kenya. *Agriculture & Food Security*, *4*(1), 8. http://doi.org/10.1186/s40066-015-0028-2

Palermo, E. (2015). Drones Could Grow to $11 Billion Industry by 2024. Retrieved August 3, 2015, from http://www.livescience.com/47071-drone-industry-spending-report.html

Porter, J. R., & Semenov, M. a. (2005). Crop responses to climatic variation. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *360*(1463), 2021–2035. http://doi.org/10.1098/rstb.2005.1752

Prasad, A. K., Chai, L., Singh, R. P., & Kafatos, M. (2006). Crop yield estimation model for Iowa using remote sensing and surface parameters. *International Journal of Applied Earth Observation and Geoinformation*, *8*(1), 26–33. http://doi.org/10.1016/j.jag.2005.06.002

Qi, J., Chehbouni, a., Huete, a. R., Kerr, Y. H., & Sorooshian, S. (1994). A modified soil adjusted vegetation index. *Remote Sensing of Environment*, *48*(2), 119–126. http://doi.org/10.1016/0034-4257(94)90134-1

Ransom, J. (2013). Corn Growth and Management: Quick guide. Retrieved January 2, 2016, from https://www.ag.ndsu.edu/pubs/plantsci/crops/a1173.pdf

Rembold, F., Atzberger, C., Savin, I., & Rojas, O. (2013). Using low resolution satellite imagery for yield prediction and yield anomaly detection. *Remote Sensing*, *5*(4), 1704–1733. http://doi.org/10.3390/rs5041704

Reynolds, C. a., Yitayew, M., Slack, D. C., Hutchinson, C. F., Huete, a., & Petersen, M. S. (2000). Estimating crop yields and production by integrating the FAO Crop Specific Water Balance model with real-time satellite data and ground-based ancillary data. *International Journal of*

Remote Sensing, *21*(18), 3487–3508. http://doi.org/10.1080/014311600750037516

Rouse, J. W., Haas, J. A. S., & Deering, D. W. (1974). *Monitoring Vegetation Systems in the Great Plains with ERTS. Third ERTS-1 Symbosium*. Washington, DC: NASA.

Sacks, W. J., Deryng, D., Foley, J. a., & Ramankutty, N. (2010). Crop planting dates: An analysis of global patterns. *Global Ecology and Biogeography*, *19*(5), 607–620. http://doi.org/10.1111/j.1466-8238.2010.00551.x

Salami, A., Kamara, A. B., & Brixiova, Z. (2010). *Smallholder Agriculture in East Africa: Trends, Constraints and Opportunities. Working Paper No.105*.

SenseFly Ltd. (2015). Browse our camera payloads & accessories. Retrieved July 20, 2015, from https://www.sensefly.com/drones/accessories.html

Singh, R., Semwal, D. P., Rai, a., & Chhikara, R. S. (2002). Small area estimation of crop yield using remote sensing satellite data. *International Journal of Remote Sensing*, *23*(1), 49–56. http://doi.org/10.1080/01431160010014756

Smale, M., Derek, B., & Jayne, T. (2011). *Maize Revolutions in Sub-Saharan Africa* (No. Policy Research Working Paper 5659). Washington, DC.

Smale, M., & Jayne, T. (2003). Maize in Eastern and Southern Africa : “ Seeds ” of Success in Retrospect. *Food Policy*, (97), 1–79. Retrieved from http://www.ifpri.org/sites/default/files/publications/eptdp97.pdf

Stenger, B., Woodley, T., & Cipolla, R. (2009). Learning to track with multiple observers. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2647–2654). IEEE. http://doi.org/10.1109/CVPR.2009.5206634

Thornton, P. K., Jones, P. G., Alagarswamy, G., & Andresen, J. (2009). Spatial variation of crop yield response to climate change in East Africa. *Global Environmental Change*, *19*(1), 54–65. http://doi.org/10.1016/j.gloenvcha.2008.08.005

Townsend, R. F. (2015). *Ending Poverty and Hunger by 2030 An Agenda for the Global Food System* (2nd ed.). Washington, DC.

Turtoi, C., Akyildirim, O., & Petkov, P. (2012). Statistical Farm Register in the EU Acceding countries – A conceptual approach, *2012*(59), 147–161.

Väisänen, P. (2009). Integration of Registers and Survey-based Data in the Production of Agricultural and Forestry Economics Statistics.

Vallet, J., Panissod, F., Strecha, C., & Tracol, M. (2011). Photogrammetric Performance of an Ultra Light Weight Swinglet.

Van Wart, J., Kersebaum, K. C., Peng, S., Milner, M., & Cassman, K. G. (2013). Estimating crop yield potential at regional to national scales. *Field Crops Research*, *143*, 34–43. http://doi.org/10.1016/j.fcr.2012.11.018

Vermote, E., Herman, M., Morcrette, J. J., & Kotchenova, S. Y. (2006). Second Simulation of a Satellite Signal in the Solar Spectrum - Vector (6SV). *Spectrum*, *Part1*(2), 1–55.

Viña, A., Gitelson, A. A., Rundquist, D. C., Keydan, G. P., Leavitt, B., Schepers, J., … Keydan, G. (2004). Monitoring Maize (Zea Mays L.) Phenology with Remote Sensing. *Agronomy Journal*. Retrieved from http://digitalcommons.unl.edu/natrespapers

Vintrou, E., Bégué, A., Baron, C., Saad, A., Lo Seen, D., & Traoré, S. (2014). A Comparative

Study on Satellite- and Model-Based Crop Phenology in West Africa. *Remote Sensing*, *6*(2), 1367–1389. http://doi.org/10.3390/rs6021367

Vrieling, A., de Beurs, K. M., & Brown, M. E. (2011). Variability of African farming systems from phenological analysis of NDVI time series. *Climatic Change*, *109*(3-4), 455–477. http://doi.org/10.1007/s10584-011-0049-1

Vyas, S., Nigam, R., Patel, N. K., & Panigrahy, S. (2013). Extracting Regional Pattern of Wheat Sowing Dates Using Multispectral and High Temporal Observations from Indian Geostationary Satellite. *Journal of the Indian Society of Remote Sensing*, *41*(4), 855–864. http://doi.org/10.1007/s12524-013-0266-3

Wang, L., Tian, Y., Yao, X., Zhu, Y., & Cao, W. (2014). Predicting grain yield and protein content in wheat by fusing multi-sensor and multi-temporal remote-sensing images. *Field Crops Research*, *164*, 178–188. http://doi.org/10.1016/j.fcr.2014.05.001

Wang, L., Zhang, F., Jing, Y., Jiang, X., Yang, S., & Han, X. (2014). Multi-Temporal Detection of Rice Phenological Stages Using Canopy Spectrum. *Rice Science*, *21*(2), 108–115. http://doi.org/10.1016/S1672-6308(13)60170-5

Wang, M., Tao, F., & Shi, W. (2014). Corn Yield Forecasting in Northeast China Using Remotely Sensed Spectral Indices and Crop Phenology Metrics. *Journal of Integrative Agriculture*, *13*(7), 1538–1545. http://doi.org/10.1016/S2095-3119(14)60817-0

Wardlow, B. D., Egbert, S. L., & Kastens, J. H. (2007). Analysis of time-series MODIS 250??m vegetation index data for crop classification in the U.S. Central Great Plains. *Remote Sensing of Environment*, *108*(3), 290–310. http://doi.org/10.1016/j.rse.2006.11.021

Willett, J. W. (1981). Area Sampling Frames for Agriculture in Developing Countries, (May). Retrieved from http://agris.fao.org/agris-search/search.do?recordID=US201300158545

Williams, L. J., & Abdi, H. (2010). Fisher's Least Significant Difference (LSD) Test. *Encyclopedia of Research Design*, 1–6.

Woebbecke, D. ., Meyer, G. E., von Bargen, K., & Mortensen, D. (1995). Shape features for identifying young weeds using image analysis. *Transactions on American Society of Agricultural Engineering*, *38*(1), 271–281. http://doi.org/10.1016/j.compag.2010.09.013

World Bank. (2008). *Agriculture for Development*. *Agriculture* (Vol. 161). Washington, DC. Retrieved from http://econ.worldbank.org/wdr

Wu, B., & Meng, J. (2013). Remote Sensing Applications on Crop Monitoring and Prediction. *Remote Sensing of Natural Resources*, (Brown 2005), 315–332.

Yengoh, G. T. (2012). Determinants of yield differences in small-scale food crop farming systems in Cameroon, 1–17.

Zhang, J., Feng, L., & Yao, F. (2014). Improved maize cultivated area estimation over a large scale combining MODIS–EVI time series data and crop phenological information. *ISPRS Journal of Photogrammetry and Remote Sensing*, *94*, 102–113. http://doi.org/10.1016/j.isprsjprs.2014.04.023

Zurita-Milla, R., Clevers, J. G. P. W., Van Gijsel, J. a. E., & Schaepman, M. E. (2011). Using MERIS fused images for land-cover mapping and vegetation status assessment in heterogeneous landscapes. *International Journal of Remote Sensing*, *32*(November 2014), 973–991. http://doi.org/10.1080/01431160903505286

APPENDIX 1: QUESTIONNAIRE

Field level maize yield assessment in Kilosa, Tanzania
September, 2015

Cover sheet: Household information

01: Introductory statement and consent: My name is Stephen Kibet, MSc Student in University of Twente, faculty of Geo Information and Earth Observation, ITC taking Natural Resource Management course, in the Netherlands. I am currently conducting a study to assess maize yield variability at farm level using airborne imagery from Unmanned Aerial Vehicles (UAV) commonly referred to as drones, thereafter explain this variation based on farm management practices information obtained from household interviews. The study is part of my MSc academic requirement and therefore the results of the survey is purely confidential and for academic purposes only. The study partly contributes to the Spurring a Transformation for Agriculture through Remote Sensing (STARS) project study carried out at the beginning of this year in Kilosa.
The study includes both survey questions and farm visit to assess where maize was grown and delineate its boundaries. The survey will take atmost two hours.
As this is purely an academic study, there is no direct benefit you will get from the study, but the outcome of the research being a proof of concept, will be helpful in understanding whether UAV data can be used together with field data to accurately estimate maize yield at field level.
Your participation in this study is voluntary. The responses you give is anonymous and in no way will it be possible to link it back to you in the final report.
If you have any questions now or during the study, please feel free to ask. Any other questions after the study you can reach me through this email address s.kibet@student.utwente.nl

02.1 Do you have any questions about the study?

____

03: Do you agree to participate in this study?
1. Yes
2. No

04: Enter Household ID *(Household who owns farms within the two 1 by 1 km fields will be listed prior to data collection and assigned ID* ***K1HHxx*** *for household x in field 1 and* ***K2HHxx*** *for household x in field 2)*

05: Date of collection: DDMMYYYY

06: Time stamp for start of interview *(time in which the interview started as recorded automatically in the tablet)*

07: GPS coordinates of the location of the household's residence *(Automatic location recording within the inbuilt GPS in the tablet including manual/ and saving within the iPAQ)* [UTM zone 37s]

Northing	Easting

Part A: Household farm level information

A1: Name of the interviewee:

A2: Age:
a) < 17
b) 18-35
c) 36-50
d) 51-70
e) > 70

A3: Marital Status
a) Married
b) Single
c) Divorced

A4:How many household members do you have in your family *(household members refers to those who eat together under one roof)*

A5: Who are the members of your household?

Member	Number
1. Husband	
2. Wife	
3. Children	
4. Relatives	
5. Laborers	

A6: Does your household engage in any crop farming?
1. Yes
2. No

A7: Is crop farming the main economic activity of your household?
a) Yes
b) No (Specify)

A8: How many farms do you own in Kilosa within the 1km by 1 km field shown in the map? *(Indicate to the farmer boundary in the iPAQ supported by high resolution image printout map. Farm refers to separate pieces of land with separate title deed; digitize the boundaries with each having a unique ID in the table: KL1XXX for field 1 and KL2XXX for field 2)*
a) One
b) Two
c) Three
d) Four
e) More than four

A9 What is the size of each farm *(if more than one)* in acres?

A10: What is the type of ownership for each farm listed in A2 *(Populate attribute table for each farm in the tablet)?*
a) Inherited
b) Purchased
c) Gift
d) Cleared land
e) Leased
f) Squatted on land
g) Caretaker
998. Other (Specify)

A11: How many years have you owned the specified piece of land
a) < 1
b) 1-2
c) 2-3
d) 3-4
e) 4-5
f) > 5

A12: Who holds tenure of farm 04?
a) Household ownership
b) Leased/rented
c) Group owned land
d) Group leased land
e) State land
998. Other (specify)

Part B: Plot level information

B: This section relates to specific information for each plots (subdivision) within the farm)

Description	Plot number 1	2	3	etc.
B1: In each of the farms listed above (A8) *(if more than one)*, how many plots do you have within your farm *(Plot are subdivision within the main farm shown by physical boundaries such as different vegetation cover, trenches or fence, and shall be digitized within a specified farm with guidance of the farmer while field refers to the two 1 by 1 km study area: The ID will be KL1F1Pxx for field 1 farm 1 plot xx and KL2F1Pxx for field 2 farm 1 plot xx. Apart from plot ID, it will have Household ID.* a) One b) Two c) Three d) Four e) Five				
B2: What is the size of the plot (if more than 1) in acres? *(The sum acreage should add to the farm totals A9)*				
B3: How would you describe this plot, from the following list?				
a) Cropland b) Fallow c) Home garden d) Livestock field e) Woodlot 998 Other (Specify)				
B3: Crops grown in this plot for the 2015 growing season [Flag maize, if they are listed, for which there will be more detailed questions later] Value sets to be adapted to local context a) Beans b) Broccoli c) Cabbage d) Carrot e) Kale **f) Maize** g) Onion h) Peas i) Potatoes j) Rice k) Spinach l) Sugar cane m) Tomatoes n) Cassava o) Chili p) Sisal 998 Others (Specify)				
B4: What had you planted in this plot the previous season? (Value set B3; if F proceed to B5)				
B5: When did you plant maize in this plot ddmmyyyy (calendar in the tablet)				
B6: When did you harvest maize in this plot ddmmyyyy (calendar in the tablet)				
B7: What is the primary use of maize crop in this plot? a) Family consumption b) For Sale c) Exchange of goods/services d) Animal feed				
B8: During the day of planting maize in this plot was there any other vegetation cover/crops? a) Yes (If yes select from list B3 a) No (farm planted immediately after ploughing/harrowing) 998 Others (Specify)				
B9: What is currently on this plot *(tag a picture)* a) Crops listed in (B3) b) Very little vegetation cover c) Bare (ploughed/harrowed) d) Crop residue 998 Others (Specify)				
B10: During the day of planting, had it rained? *(If yes proceed to B9 and 10 otherwise only yo B8)* a) Yes b) No				
B11: How long did you plant maize in this plot before it started raining? a) < 1 Week b) 1 Week c) 2 Weeks d) 3 Weeks e) 1 month f) 1-2 months g) > 2 months				
B12: For how long did you wait to plant after the rain started (value set in B11)				
B 10: How would you describe the rainfall intensity then during planting? *(date given in B5)* a) Low b) Average c) High d) Very high				
B11: What did you consider when deciding the day to plant maize in your plot? *(Rank 1 most important and 5 less important)*				

Factor	Rank
a) Availability of funds	

b) Other farmers
c) Rainfall availability
d) Inputs availability
e) When you like

B12: Which household member has final say about the following decisions regarding this plot (Value set A5)?
a) Acquiring or releasing the plot
b) Which crops to grow?
c) How to cultivate the crops? (e.g., cropping systems, management practices, timing, etc.)
d) What to do with the products? (e.g., own use, sale, giving away, etc.)

B13: Which household members provided the primary labor for this plot? (Multiple choice value set A5)

B14: What is the gradient of the majority of the plot? *(Elevation map/physical check)*
a. Flat (<10°)
b. Gently sloping (10°-45°)
c. Steep (>45°)

B15: What is the soil type (*the most dominant soil type on the plot based on touch and feel*)?
a. Mostly sand
b. Mostly clay
c. Mostly silt
d. Loam (composed mostly of sand and silt, and a smaller amount of clay)
e. Sandy-clay
f. Silty-clay
g. Rocky
998. Other
999. Don't know

B16: What is the soil fertility?
a) Poor
b) Fair
c) Good
d) Very good

Part C: Maize production and management information

This refers to maize production and manage during the 2015 growing season.

	Plot number = 1				2
C1: What types of strategies did you use to water (*tick all that apply*) in each growing season? a) Rain fed b) Drip irrigation c) Flood irrigation d) Border irrigation e) Furrow irrigation f) Sprinkler irrigation					
C2: How many maize bags did you harvest in this plot (*specified by the ID*) during the 2015 growing season *(Threshed maize bag, if not threshed estimate how many bags will make one threshed bag?)*					
C3: How many maize bags did you harvest the previous season?					

C4: How has land productivity (in terms of Maize yields) changed in the past five years?
a. Decreased with increased inputs
b. Decreased with same inputs
c. Decreased with decreased inputs
d. Increased with increased inputs
e. Increased with same inputs
f. Increased with decreased inputs
g. Unchanged with increased inputs
h. Unchanged with same inputs
i. Unchanged with decreased inputs

Part D: Maize crop management

D1: Did you adopt any of these cropping systems for maize during the growing season? *(Only 1 option possible)*
a) Monocropping (Maize was cultivated without mixing with other crops)
b) Intercropping (Maize was intercropped with other crops)
c) Mixed cropping (Maize was mixed-cropped with other crops)

D2: For the answer in D1 select the crops intercropped or mixed with maize from value set B3

D3: How was the land prepared for maize before the start of the season? (select multiple)
a) Hand hoe
b) Oxen
c) Machinery
d) Herbicide
e) Don't do any preparation
f) 998 Other

D4: Which of these chemical fertilizers did you use on this crop in each growing season? (select all that apply)
a) NPK 17 17 17
b) CAN 26 0 0
c) UREA 46 0 0
d) DAP 18 46 0
e) MOP 0 0 60

D5: For each of the chemical fertilizers selected above, please answer the following for each growing season:

Type	Unit (if answered "bag", ask how much a bag weighs)	Unit price	Quantity applied
a) NPK 17 17 17			
b) CAN 26 0 0			
c) UREA 46 0 0			
d) DAP 18 46 0			
e) MOP 0 0 60			

D10: Which of these organic fertilizers did you use on this maize field during the growing season? (select all that apply)
a. Animal manure
b. Compost
c. Crop residues or other dead plant
d. Other (specify)

D11: For each of the chemical fertilizers selected above, kindly provide this information growing season:

Type	**Unit (if answered "bag", please ask how much a bag weighs)**	**Unit price**	**Quantity applied**
e. Animal manure			
f. Compost			
g. Crop residues or other dead plant			
h. Other (specify)			

Insect pest management

D12: Can you identify location within the farm where insect pest attacked your maize crops? (GPS locations within the plot)

E13: What was the most significant insect pest for this crop in this location during the growing season? (select one) [Adapted per field with photos to aid description)
a. Stalk borer
b. Cutworms
c. Thrips
d. Aphids
e. Diamond moths
998 Other

D14: How would you describe the **extent of maize damage** from this insect pest without control for maize in this location of the plot?
a) Not a problem – little to no crop loss
b) Inconvenience – minor crop loss
c) Mild problem– moderate crop loss
d) Large problem – large crop loss
e) Catastrophic – significant crop loss

D15: To control this insect pest, what was the primary method you applied to control this insect pest for maize in this section of the plot (select one)
a) Chemical insecticide application
b) Rely on predator insects
c) Pest-resistant crop varieties
d) Cultural control (rotation, intercropping, planting times, etc.)
e) Field burning
f) Natural or plant-based repellents
g) Mechanical/manual control (hand picking)
h) Irrigation (washing off)
i) Weather (eg., rainfall)
j) Don't take any action
998 Other (specify)

D16: In addition to the primary control method that you selected above, did you use any of the other control methods to control this insect pest for maize during the growing season? (select all that apply)
a) Chemical insecticide application
b) Rely on predator insects
c) Pest/disease-resistant varieties
d) Cultural control (crop rotation, intercropping, planting times etc)
e) Field burning
f) Natural or plant-based repellents
g) Mechanical/manual control (hand picking
h) Irrigation (washing off)
i) Weather (rainfall)
j) Don't take any action
998 Other (Specify)

Disease management

D19: Can you identify location within the farm where disease attacked your maize crops? (GPS locations within the plot)

D20: What was the most significant plant disease for maize in this location during the growing season? (select one) [ADAPT TO LOCAL CONTEXT]
a) Maize Streak Virus
b) Maize lethal Necrosis

D21: How would you describe the extent of crop damage from this disease without control for maize during the growing season?
a) Not a problem – little to no crop loss
b) Inconvenience – minor crop loss
c) Mild problem– moderate crop loss
d) Large problem – large crop loss
e) Catastrophic – significant crop loss

D22: To control this plant disease, what was the primary method you applied to control this disease for maize during the growing season? (select one) [Adapt to local context]
1) Chemical disease application
2) Disease-resistant crop varieties
3) Cultural control (sanitation, crop rotation, host eradication, improvement of crop environment)
4) Physical method (e.g., removal of plants)
5) Establish barrier
6) Field burning
7) Irrigation (washing off)
8) Weather (eg., rainfall)
9) Don't take any action
998 Other (specify)

D23: In addition to the primary control method that you selected above, did you use any of the other control methods to control maize disease during the growing season? (select all that apply)
a) Chemical disease application
b) Pest/disease-resistant varieties
c) Cultural control (crop rotation, intercropping, planting times etc)
d) Field burning
e) Natural or plant-based repellents
f) Mechanical/manual control (hand picking
g) Weather (rainfall)
h) Don't take any action
998Other (Specify)

D24: How effective did you find this method in controlling this disease? (applying to whichever control method selected)
a) Not effective (disease intensity reduction less than 10%)
b) Somewhat effective (disease intensity reduced by 11%- 40%)
c) Effective (disease intensity reduced by 41%- 80%)
d) Very effective (eliminate >80% of disease)

D25: How many times did you repeat this method (i.e., total number of applications) during the maize growing season? (applying to control methods 1, 4, 6, 7, if selected)

D26: On average, how much family labor was used for **each application** (person-days)? (applying to control methods 1, 4, 6, 7, if selected)

D27: On average, how much hired labor was used for **each application** (person-days)? (applying to control methods 1, 4, 6, 7, if selected)

D28: Which household member was mainly responsible for undertaking this plant disease control activity (applying to control methods 1, 4, 6, 7, if selected)? (Select from value set B3)

D29: *If applying chemical pesticides (selected option 1 in question D22):* What was the name of the primary pesticide product you purchased to treat this plant disease in each growing season? (type names)

D30: What was the main equipment used to spray this pesticide?

1) Hand sprayer
2) Knapsack sprayer
3) Automatic sprayer
4) Boom sprayer
5) Don't know
6) Other

D31: How much did this disease control product cost per unit? (TSH per unit)

D32: What was the purchased unit (g, kg, ml or l)?

D33: When spraying against this maize disease, did you mix this product with any insecticides?

a) Yes
b) No

D34: When spraying against this maize disease, did you mix this c product with any herbicides?

d. Yes
e. No

Weed management

D35: What was the primary weeding for maize during the growing season? (select all that apply)

a) Burning
b) Crop rotation
c) Did not weed
d) Herbicide (before planting)
e) Herbicide after planting)
f) Manual pulling
g) Manual weeding with hoes
h) Weeding with oxen
i) Weeding does not matter
998) Other (please specify)

D36: What are the common weeds that was in your maize farm (type)

D37: How many times did you repeat this method (i.e., total number of applications) during the maize growing season? (applying to control methods d, e, g, h, if selected)

D38: During each of the weeding listed in D37 did you carry out for the:-

a) Entire plot
b) Sections of the plot (GPS coordinates)

D39: How effective did you find this method in controlling these weeds? (applying to whichever control method selected)

a) Not effective (weed intensity reduction less than 10%)
b) Somewhat effective (weed intensity reduced by 11%- 40%)
c) Effective (weed intensity reduced by 41%- 80%)
d) Very effective (eliminate >80% of weed)

D40: On average, how much family labor was used for **each application** (person-days)? (applying to control methods d, e, g, h, if selected)

D41: On average, how much hired labor was used for **each application** (person-days)? (applying to control methods d, e, g, h, if selected)

D42: Which household member was mainly responsible for undertaking this plant weeding control (as selected in value set B3)

D43: *If applying chemical herbicide (selected option d, e in question E31):* What was the name of the primary herbicide product you purchased to control weed during the growing season? (type names)

D44: What was the main equipment used to spray this pesticide?

7) Hand sprayer
8) Knapsack sprayer
9) Automatic sprayer
10) Boom sprayer
11) Don't know
998Other

E45: How much did this insecticide product cost per unit? (TSH per unit)

E46: What was the purchased unit (g, kg, ml or l)?

Part E: Shocks

Did you experience the following shocks during the maize growing season 2015*? (Shocks are events that caused significant maize losses to the household)*

			For the top three shocks only		
Shock	Severely affected by this shock?	Rank the three most severe shocks	This [SHOCK] primarily affected…	How long ago did this shock occur? Years	Months
Drought/Floods					
Crop disease or crop pests					
Large decrease in maize prices					
Large increase in farm input prices (e.g., fertilizer)					
Severe water shortage					
Other					
Code	1. No 2. Yes	Rank 1-3 1 most severe 3 less severe)	1. Own HH only 2. Some but not all HH in village 3. All HH in village	Years	Months

Part F: Conclusion:

Thank you so much for spending time with me to answer these questions.