

A comparative analysis of PlanetScope and Sentinel-2 space-borne sensors in mapping Striga weed using Guided Regularised Random Forest classification ensemble.

B. T. Mudereri^{1,2*}, T. Dube², E. M. Adel-Rahman^{1,3}, S. Niassy¹, E. Kimathi¹, Z. Khan¹, T. Landmann⁴

¹ International Center of Insect Physiology and Ecology (ICIPE), P.O. Box 30772, 00100 Nairobi, Kenya - (*bmudereri, eabdel-Rahman, sniassy, ekimathi, zkhan)@icipe.org

² Department of Earth Sciences, University of Western Cape, Private Bag X17 Bellville 7535 South Africa - tidube@uwc.ac.za

³ Department of Agronomy, Faculty of Agriculture, University of Khartoum, Khartoum North 13314, Sudan

⁴ RSS-Remote Sensing Solutions Gmbh, Dingolfinger Str. 9, 81673 Munich, Germany - Landmann@rssgmbh.de

KEYWORDS: Feature selection; Food security; High spatial-temporal resolution; Nanosatellites; New generation sensors; Precision agriculture; Weed detection; Sub-Saharan Africa.

ABSTRACT:

Weeds are one of the major restrictions to sustaining crop productivity. Weeds often outcompete crops for nutrients, soil moisture, solar radiation, space and provide platforms for breeding of pests and diseases. The ever-growing global food insecurity triggers the need for spatially explicit innovative geospatial technologies that can deliver timely detection of weeds within agro-ecological systems. This will help pinpoint maize fields to be prioritized for weed control. Satellite remote sensing offers incomparable opportunities for precision agriculture, ecological applications and vegetation characterisation, with vast socioeconomic benefits. This work compares and evaluates the strength of Sentinel-2 (S2) satellite with the constellation of Dove nanosatellites i.e. PlanetScope (PS) data in detecting and mapping Striga (*Striga hermonthica*) weed within intercropped maize fields in Rongo sub-county in western Kenya. We applied the S2 and PS derived spectral data and vegetation indices in mapping the Striga occurrence. Data analysis was implemented, using the Guided Regularised Random Forest (GRRF) classifier. Comparatively, Sentinel-2 demonstrated slightly lower Striga detection capacity than PlanetScope, with an overall accuracy of 88% and 92%, respectively. The results further showed that the VNIR (Blue, Green Red and NIR) and the Atmospheric resistance Vegetation Index (ARVI) were the most fundamental variables in detecting and mapping Striga presence in maize fields. Findings from this work demonstrate that Sentinel-2 data has the capability to provide spatial explicit near real-time field level Striga detection – a previously daunting task with broadband multispectral sensors.

1. INTRODUCTION

1.1 Background

Timely and repeatable spatial explicit information on inter and intra-field variability of agro-ecological systems is key to devise adaptive and informed management decisions regarding crop productivity, resource allocation and labour efficiency (Houborg and McCabe, 2018a). Advancements in remote sensing technologies, data analytics, research and monitoring initiatives of invasive pests and weeds provide tremendous benefits to the continuous and near real time observations of crop health and crop phenological development (Mutanga et al., 2017). However, there is an inevitable unbalanced trade-off between the spatial resolution and frequency of observations that can be derived from the conventional satellites. Generally, most conventional satellites providing very frequent observations (daily and near daily revisit time) lack in spatial resolution, such as the Moderate resolution Imaging Spectrometer (MODIS) (Masocha et al., 2018). High frequency of observation coupled with high spatial resolutions facilitates the rapid detection evolving from crop dynamics at spatial scales, high enough for timely interventions and effective resource management (Houborg and McCabe, 2018a).

The advent of the CubeSat (~ 10 Kg) in particular, data from Planets Labs Inc, provides satellite images collected by a constellation of nanosatellites (Doves) at very high spatial resolutions (VHSR) (Baloloy et al., 2018). Constellations of these 175 CubeSats are proving to be an innovative source of data, with vast potential to overcome the spatial-temporal limitations of conventional single sensor satellite missions (Houborg and McCabe, 2018b).

The aforementioned nanosatellites are equipped with unique multispectral sensors VNIR (420-900nm) at 3m spatial resolution and provide daily global data equivalent to a daily collection capacity of 350 million km²/day (Planet, 2018). However, since the first launch of the first flock of Doves on the 22nd of June 2016, research on the possible benefits afforded by PlanetScope data for landscape species characterisation is still limited (Cooley et al., 2017). This could be attributed to the newness of CubeSats, the cost of imagery, concerns over cross-sensor calibration, location accuracy, image quality, data availability and accessibility (Cooley et al., 2017). However, these nanosatellites can essentially improve spatial analysis methods for weed identification, detection and mapping within agro-ecological systems by providing consistent information showing in-field variability in weeds configuration and growth patterns (Yue et al., 2018).

Although the CubeSats are superior in their spatiotemporal resolution, the radiometric coverage is inferior to the conventional satellites such as Sentinel-2 (S2). This research evaluates and intercompares the capabilities and potential scientific utilities of S2 data with the very high spatial resolution nanosatellite data of PlanetScope (PS) to detect and map Striga (*Striga hermonthica*) in intercropped African farm fields. S2 is a combination of two sensors (2A-2B) developed by European Space Agency (ESA) (Chemura et al., 2018). These sensors provide images at 10, 20 and 60 m spatial resolution and a spectral range of 440 to 2280 nm at a combined global average revisit time of 5 days (Immitzer et al., 2016). Although several capabilities of S2 have been successfully tested, such as evaluating rangeland quality (Ramoelo et al., 2015; Shoko and Mutanga, 2017), above biomass estimation (Sibanda et al., 2015)

water hyacinth mapping (Thamaga and Dube, 2018) and understanding crop behaviour (Veloso et al., 2017; Vrieling et al., 2018) its utility in detecting and mapping *Striga* occurrence is not documented. The use of S2 imagery is advantageous because of its spatiotemporal coverage, the presence of the strategically positioned Red-edge bands, 13 multispectral bands, its global foot-print and free-availability.

Striga is one of the severe biological crop pests reducing food production in Sub Saharan Africa resulting in more than US\$1billion in losses every year (Atera et al., 2013; Scholes and Press, 2008). *Striga* is a parasitic weed whose existence relies on availability of the cereal crops in particular maize and sorghum (Khan et al., 2002; Midega et al., 2017). Since most optical remote sensing instruments capture top of canopy reflection, the use of spaceborne sensors in mapping *Striga* occurrence pose huge mapping challenges as the weeds mostly occur underneath cereal plants. Crop health condition and phenology metrics entrenched within vegetation indices are, therefore, crucial to predict and inform on the existence of the understory weed.

In this study, we evaluated the strength and usability of S2 compared to the PS data in detecting and mapping *Striga* occurrence within maize fields in western Kenya, using GRRF classification ensemble. GRRF algorithm is superior in its ability to identify key variables and robust in data processing capabilities (Mureriwa et al., 2016). GRRF is regarded as one of the most robust and effective classification method for agro-ecological systems, using very high spatial resolution satellites like PS, particularly within heterogeneous landscapes (Adam et al., 2017; Kyalo et al., 2017).

2. MATERIALS AND METHODS

2.1 Study area

The study was conducted in Rongo Subcounty, Migori County of Western Kenya. Rongo is administratively divided into four wards namely; North, East, Central and South Kamagambo (Figure 1). The town is located at latitude -0.75786°S and longitude 34.60901°E at an altitude of 1470m. The study area is bound by the coordinates $0^{\circ} 39' 12'' \text{ S}$; $34^{\circ} 35' .40'' \text{ E}$ and $0^{\circ} 59' 16'' \text{ S}$; $34^{\circ} 37' 21'' \text{ E}$. It receives an average annual rainfall of 1833 mm, with a bimodal yearly seasonal pattern. Average annual temperature is around 22°C , and humidity ranges between 50 and 70 %.

The agro-ecological system is characterized by loam, sandy and clay soils, with most of the agricultural activities being subsistence and small-scale agriculture. Crops grown in Rongo include maize, beans, groundnuts, green grams, cassava and some horticultural crops such as mangoes, banana, avocado, pawpaw and indigenous vegetables with sugarcane being the main cash crop in the area. The crop fields vary ($\pm 30 \text{ m}^2$) in dimension. Due to the limited farmlands allocated to individual farmers, intercropping of maize with either beans, cowpea, and groundnuts is the main farming practice in the area.

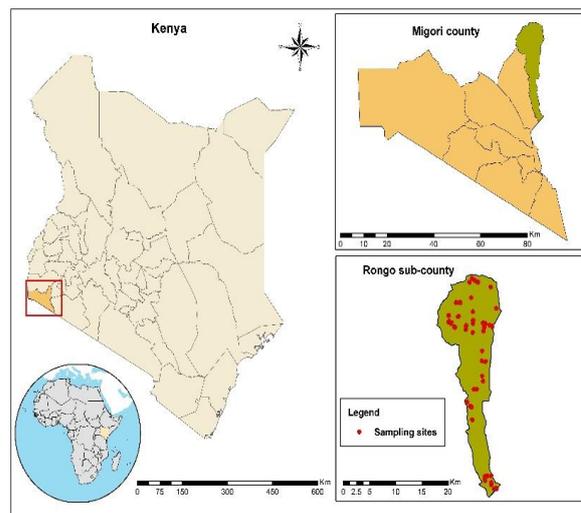


Figure 1: Domain and geographic location of the study area in Kenya and Migori subcounty.

2.2 Data collection

Field surveys were conducted between the 12th to the 16th of December 2017. This period coincided with peak *Striga* flowering phase and maize reproductive stages R1-R6. Reference data were gathered within croplands. In our study area, cropland system was mainly mono and mixed maize cropping. A stratified random sampling approach, using the administrative ward boundaries as strata was followed to determine areas with *Striga* the presence or absence reference data. A handheld Global Positioning System (GPS) instrument with an accuracy margin of $\pm 3 \text{ m}$ was used to locate the reference data. We collected the presence or absence data at the centre of the field to avoid the influence of edge effect. We geo-tagged each plot in the sampled fields using photographs taken from the centre and the four cardinal directions for supplementary analysis of the cropping systems and crop age. Built-up area, Bare patches, Water and Natural vegetation reference data were collected by digitising Regions Of Interest (ROI) using visual interpretations of very high spatial resolution imagery within Google Earth® as described by Chemura et al., (2017) and summarised in Table 1.

Class	Class ID	Description
Bare patches	1	Surfaces without vegetation
Built-up	2	Human made constructions
Natural vegetation	3	Wood vegetation and grasslands
Non- <i>Striga</i> fields	4	Crop fields without <i>Striga</i>
<i>Striga</i> infested fields	5	Crop fields infested with <i>Striga</i>
Water	6	Water bodies

Table 1: Landcover classes used in the classification analysis for both PlanetScope and Sentinel-2 images

2.3 Image acquisition and pre-processing

Images of PS (16th December 2017) and S2 (13th December 2017) were used in the analysis. PS was provided by Planet labs Inc <https://www.planet.com> and delivered as analytic 4-band product (VNIR) (Planet, 2018). The data was received as an Ortho Scene product (level 3B), which relates to Top of Atmosphere (TOA) radiance suitable for analytic and visual application. The scenes were provided already orthorectified to $<10\text{m}$ RMSE position accuracy and projected to UTM/WGS84 cartographic projection. The scaled radiance was converted to TOA reflectance using a Planet Labs python guide

(<https://developers.planet.com/tutorials/convert-planet-scope-imagery-from-radiance-to-reflectance/>) and the corresponding reflectance calibration coefficients associated with each band from the metadata.xml files. The scenes were mosaiced to cover the entire study area. Similarly, S2 data, processing level 1C was freely downloaded from the Copernicus data download platform <https://scihub.copernicus.eu/dhus/#/home> within the granule T36MXE. Level 1C data from Sentinels are provided as TOA reflectance, already orthorectified in cartographic geometry in tiles of 100 km², UTM/WGS84 projection. These were converted to atmospherically corrected, surface reflectance Level 2A using the SEN2COR in SNAP v6.0 software run using the default parameter settings. The choice of images was informed by the alignment with the same period the field reference data was collected, the availability from the sensor's archive and the corresponding cloud cover. All the images were acquired on days of low cloud cover (<5%) and during the period when vegetation was in full vigour.

2.4 Broadband vegetation indices

Seven vegetation indices were computed from PS and S2 images (Table 2). Visible and Near-infrared bands were used for this study to derive vegetation indices. The indices were computed, using the freely available SNAP v6.0 software. These indices were chosen founded on (1) their ability to reduce soil background effects, (2) their strength to separate between flowering and nonflowering plants and (3) capability to enhance vegetation greenness (Xue and Su, 2017). The Normalized Difference Vegetation Index (NDVI) has been successfully used to estimate biomass and crop yields (Chemura et al., 2017; Dube et al., 2015). Also, the usage of NDVI as opposed to the classification of raw bands is recommended for discrimination of vegetation from other non-photosynthetic classes (Bannari et al., 1995; Matongera et al., 2017). Ratio Vegetation index (RVI) is based on the principle that leaves absorb more red than Near Infrared (NIR) (Bannari et al., 1995; Xue and Su, 2017). Two indices that correct for the influence of the soil background were used namely; the Modified Secondary Soil adjusted Vegetation Index (MSAVI2) and the Soil Adjusted Vegetation Index (SAVI). MSAVI2 does not rely on the soil line to eliminate the soil influence but introduces a function rather than a constant “L” value, which is secondary modification of SAVI and MSAVI. SAVI minimises the influence of soil brightness by introducing the soil condition “L”. The value of L ranges between 0 to 1, where values of L are close to zero, SAVI equals NDVI (Royimani et al., 2019). In this study we used L = 0.5 which is common practice for most environmental conditions (Xue and Su, 2017). The Atmospherically Resistant Vegetation Index (ARVI) is based on the correction of the atmospheric influences through the elimination of the effects of atmospheric aerosols using the difference between the blue (B) and red (R) bands. The Difference Vegetation Index (DVI), which subtracts the red band from the near infrared band, has been useful for vegetation monitoring, however, it is very sensitive to changes in the soil background (Bannari et al., 1995). Infrared Percentage Vegetation Index (IPVI) is recommended for vegetation mapping as it is mostly sensitive to vegetation cover and vegetation biomass (Royimani et al., 2019)

Vegetation index	Equation	Reference
Normalised Difference Vegetation Index (NDVI)	$\frac{NIR - R}{NIR + R}$	(Abdel-Rahman et al., 2016)
Infrared Percentage Vegetation index (IPVI)	$\frac{NIR}{NIR + R}$	(Royimani et al., 2019)
Simple Ratio/ Ratio Vegetation Index (RVI)	$\frac{NIR}{R}$	(Evangelista et al., 2009)
Soil Adjusted Vegetation Index (SAVI)	$\frac{NIR - R}{NIR + R + L} (1 + L)$	(Hadjimitsis et al., 2010)
Modified Secondary Soil Adjusted Vegetation index (MSAVI2)	$0.5 \times [2 NIR + 1 - \sqrt{2(NIR + 1)^2 - 8(NIR - R)}]$	(Xue and Su, 2017)
Atmospheric Resistant Vegetation Index (ARVI)	$\frac{NIR - RB}{NIR + RB}$	(Xue and Su, 2017)
Difference Vegetation Index (DVI)	$NIR - R$	(Dube et al., 2015)

Table 2: Selected vegetation indices for the discrimination of Striga infested crop fields from other landcover classes: where RB in the Atmospheric Resistance Vegetation Index is the difference between the red band and blue band

2.5 Guided Regularised Random Forest

We used Guided Regularized Random Forest (GRRF) to select the most important bands and indices for the prediction of the six landcover classes (Table 1). Many researchers have used random forest as a dimensionality reduction tool to reduce data redundancy within explanatory variables (Dube et al., 2014; Han et al., 2017). However, research has shown disadvantages for the use of random forest as a tool to measure variable importance, as well as a variable selection method (Murériwa et al., 2016). The package “RRF” in R software was used to perform the variable importance analysis and classification (Deng, 2013; R Core Team, 2018). A *gamma* value of 0.6 was used in the “CoefReg” function to determine the limit used to restrain the explanatory variables to the minimum without compromising on the capacity of the algorithm to predict accurately. Compared to the standard algorithms, the regularized framework significantly reduces the training time by building a single model (Deng and Runger, 2013). The GRRF uses a similar concept of a RF model but uses the importance scores generated from RF to guide the variable selection process (Adam et al., 2017; Murériwa et al., 2016). The importance value of a variable in RF is attained through the “Gini index” over all nodes, across all trees generated and the feature is used to split measure the pureness of the feature at every node to facilitate the voting process (Breiman, 2002). Hence, GRRF provides the precise variables that are most suitable for predicting the feature, out of the multiple features (Deng, 2013). Like the

standard algorithm of random forest for classification, the Regularised Random Forest algorithm is an ensemble machine learning approach, which combines a large set of decision trees (Breiman, 2001). The approach provides randomness by the bagging operation to fit numerous decision trees on random subsamples (Abdel-Rahman et al., 2013a; Adam et al., 2014). The multiple classification trees then vote by plurality on the correct classification (Royimani et al., 2019). Most studies rely on the Out of Bag (OOB) accuracy for the model evaluation, however the current study used a 10 times X 10-fold repeated cross-validation technique, set to train and validate the performance of the models (Immitzer et al., 2016). The three parameters, *mtry* and *ntree* = 500, *flagReg* = 1, were used as default using the tune length of 3. According to (Abdel-Rahman et al., 2013b) increasing *ntree* beyond 500 does not show a pronounced effect on Root Mean Square Error (RMSE).

2.6 Accuracy Assessment

Data generated from the digitised training areas were split into two datasets 70% and 30%, which is “Training” and “Testing” respectively, as per the standard machine learning evaluation criteria (Dube et al., 2014; Kyalo et al., 2017). Model testing data was used in the construction of the confusion matrices for the validation of the performance of the GRRF model. The Overall Accuracy (OA), User’s Accuracy (UA) and the Producer’s Accuracy (PA) were used together with Kappa statistics to evaluate the performance of the classifier and the sensors.

3. RESULTS AND DISCUSSION

3.1 Variable selection using Guided Regularised Random Forest

Results from both PS and S2 imagery show that the visible bands (Blue, Green and Red) were critical for separating the two Striga classes (Striga and non-Striga infested fields) from the four land cover classes (bare patches, built-up, natural vegetation and water). The GRRF was able to determine only six key variables from S2 and five variables for prediction using PS. Amongst the vegetation indices, only ARVI and NDVI for S2 were selected by the GRRF whilst ARVI and IPVI were selected for the PS (Figure 2). These results prove the strength and importance of the raw bands of these two sensors in discriminating vegetation from manmade features. However, the indices were also crucial for highlighting plant inherent properties associated with leaf properties, such as chlorophyll content and vigour. Vegetation indices have specific expressions and ratios, which can represent green vegetation properties better than using individual bands (Baloloy et al., 2018). NDVI has been tested in various related research and has proven a valuable index correlated to biophysical plant characteristics such as chlorophyll content and LAI (Baugh and Groeneveld, 2006). Similarly, ARVI is a modified version of NDVI that overcomes the influence of the atmosphere.

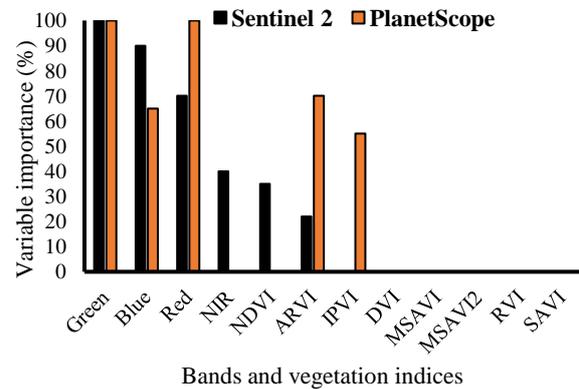


Figure 2: Variable importance percentage for both Sentinel-2 and PlanetScope derived metrics using the GRRF variable selection measure

3.2 Mapping accuracy assessment

The use of S2 selected variables (Red, Green, Blue, Near-infrared and ARVI) produced an overall accuracy of 87% and Kappa of 0.82 in detecting Striga occurrence in maize fields (Figure 3). The results were slightly lower (-5% deviation) compared to those obtained using PS. The results show that PS selected variables (Red, Green, Blue, ARVI and IPVI) produced the slightly higher Land Use/Land cover (LULC) mapping results (92%) and Kappa of 0.89. The use of a few systematically selected number of effective bands has shown that it is possible to attain and exceed the classification accuracy of the entire waveband dataset (Cao et al., 2018). This also reduces the redundancy produced by correlated variables. The GRRF process decreases the multidimensionality of the variable data without compromising key information relevant to the features (Adam et al., 2017; Deng, 2013; Mureriwa et al., 2016)

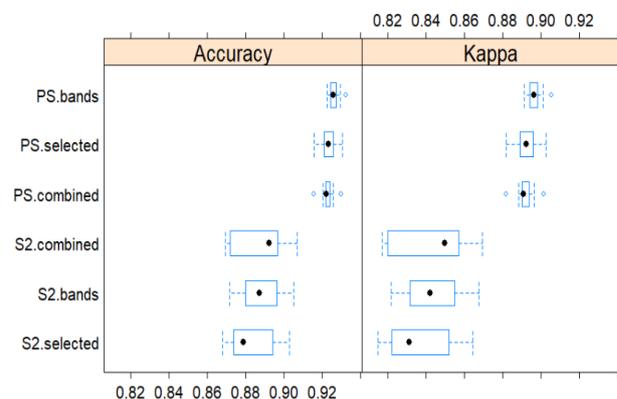


Figure 3: Model prediction performances using Accuracy and Kappa for bands only variables (*.bands), combined indices and bands (*.combined) and Guided Regularized Random Forest (GRRF) selected variables (*.selected) using PlanetScope (PS) and Sentinel-2 (S2) imagery.

Although S2 and PS both performed very well in predicting other classes, the accuracy results for Striga and non-Striga fields was low 66% and 72% respectively (Table 3). Based on the accuracy produced by the single scene of PS, these results demonstrate the ±175 CubeSats constellation are a relevant source of spatial data capable of consistently monitoring the presence of Striga and other landcover classes. However, due to the limiting costs of acquisition of the PS imagery, S2 produced acceptable accuracies, which are essential for monitoring Striga occurrence.

These classification errors are related to the split of cropland classes into Striga and non-Striga crop fields, which have similar spectral responses. The split was necessary to differentiate maize fields infested with Striga and non-infested fields. The classification error margins could also be attributed to the complexity produced by the miniature (< 30m²) heterogeneous fields mainly with intercropped plants which characterise the study area. Since other non-cropland classes could be classified with very high accuracy these results could potentially be improved by masking other classes except cropland and predicting Striga within the cropland class.

Class	PlanetScope				Sentinel-2			
	Combined variables		GRRF selected variables		Combined variables		GRRF selected variables	
	PA	UA	PA	UA	PA	UA	PA	UA
Bare	92	95	93	96	88	92	88	91
Built-up	93	87	93	88	92	85	90	85
Natural Vegetation	98	99	98	99	96	95	94	96
Non Striga Fields	70	70	67	70	51	52	53	56
Striga Fields	65	72	67	68	53	66	54	52
Water	89	94	91	91	100	97	93	93
OA	92%		92%		88%		87%	
Kappa	0.89		0.89		0.83		0.82	

Table 3: Summarized confusion matrices and classification accuracies, Overall accuracy (OA), Producer’s Accuracy (PA) User’s Accuracy (UA) and Kappa statistics using the Guided Regularised Random Forest image classification of PS and S2 comparing combined variables (all bands and all indices) and GRRF selected variables (fewer optimum bands and indices selected with the Guided Regularised Random Forest algorithm).

3.3 Striga mapping

The final thematic maps produced via the GRRF algorithm are shown in Figure 4. Both images show that much of the Striga infestation is in the North and North-West of Rongo whilst low on the south. The south is mainly characterised by sugarcane farms, which are less susceptible to Striga infestation whereas the North and Northwest regions are predominantly maize fields. There are patches of Striga infested fields within the built-up area surrounding Rongo Town. Most of the urban dwellers practice urban agriculture, hence the presence of Striga within the vicinity of the central business district.

Both images ably predicted for the built-up, water, natural vegetation and bare areas. The ability of these two sensors and the capability of the GRRF model to separate these classes is commendable. Interestingly, there were able to classify the river quite distinctively from the rest of all the surrounding classes. Apart from the capacity to map striga these results confirm the potential of S2 to monitor LCLU changes over large landscape scales.

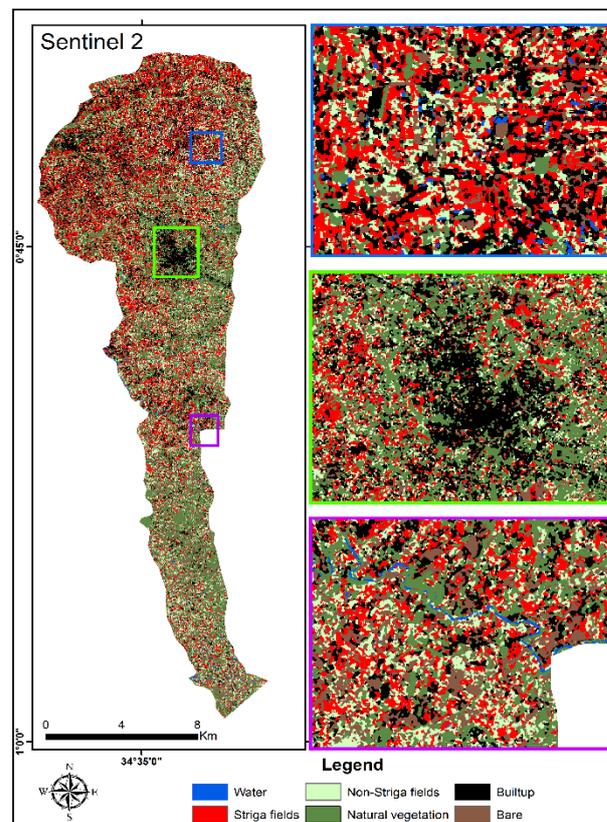
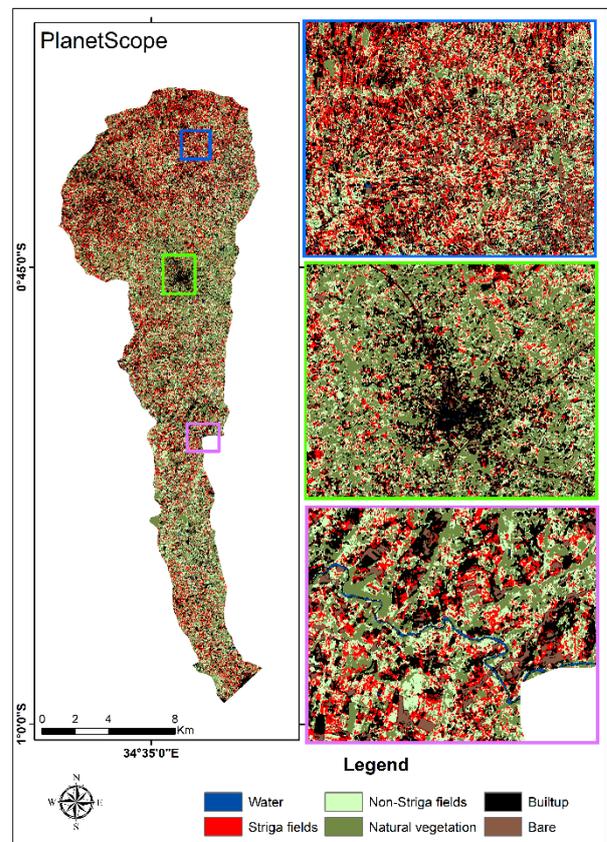


Figure 4: Land use and land cover map of Rongo showing the distribution of Striga infested fields from the GRRF classification of PlanetScope and Sentinel-2.

4. CONCLUSION

We conclude that:

- Sentinel 2 sensor was capable of detecting and mapping Striga infested maize fields with high accuracy, almost similar those derived using PS metrics.
- The GRRF feature selection shortlisted the Red, Green, Blue, Near-infrared and ARVI as the most important variables for predicting Striga occurrence, with an overall classification accuracy of 87% (-5% deviation from PS selected variable performance).

Overall, the findings of this work provide baseline information appropriate to devise adaptive weed management strategies in Sub-Saharan Africa. However, since the results proved that there were errors of commission between Striga and non-Striga fields caused by similarities in the spectral behaviours of the crops, we recommend that a hierarchical approach be tested by masking out croplands and using subpixel-unmixing technics to improve the detection ability of Striga occurrence. Since Striga swiftly invades agro-ecological landscapes, we also suggest landscape scale monitoring of Striga using multi-date Sentinel 2 data. Long-term remote sensing applications are effective in defining spatial trends and evolution of the distribution invasive weeds over time

5. ACKNOWLEDGEMENTS

The authors acknowledge with appreciation the financial support received from Biovision Foundation for Ecological Development (Switzerland), the German Academic Exchange Service (DAAD) and the International Centre of Insect Physiology and Ecology (*icipe*). We are also thankful to the collaboration work with Keele University (UK) and for providing the PlanetScope data. In the same way, the authors express their gratitude to the farmers in Rongo for their contribution through information sharing and cooperation. We also extend our gratitude to Mr Martin Ogola Oluoch and Mr Kennedy Okeyo Anyango for their support in the field.

6. REFERENCES

Abdel-Rahman, E.M., Ahmed, F.B., Ismail, R., 2013a. Random forest regression and spectral band selection for estimating sugarcane leaf nitrogen concentration using EO-1 Hyperion hyperspectral data. *Int. J. Remote Sens.* 34, 712–728. <https://doi.org/10.1080/01431161.2012.713142>

Abdel-Rahman, E.M., Landmann, T., Kyalo, R., Ong'amo, G., Mwalusepo, S., Sulieman, S., Ru, B. Le, 2016. Predicting stem borer density in maize using RapidEye data and generalized linear models. *Int. J. Appl. Earth Obs. Geoinf.* 57, 61–74. <https://doi.org/10.1016/j.jag.2016.12.008>

Abdel-Rahman, E.M., Way, M., Ahmed, F., Ismail, R., Adam, E., 2013b. Estimation of thrips (*Fulmekiola serrata* Kobus) density in sugarcane using leaf-level hyperspectral data. *South African J. Plant Soil* 30, 91–96. <https://doi.org/10.1080/02571862.2013.803616>

Adam, E., Deng, H., Odindi, J., Abdel-Rahman, E.M., Mutanga, O., 2017. Detecting the early stage of phaeosphaeria leaf spot infestations in maize crop using in situ hyperspectral data and guided regularized random forest algorithm. *J. Spectrosc.* 2017. <https://doi.org/10.1155/2017/6961387>

Adam, E., Mutanga, O., Abdel-Rahman, E.M., Ismail, R., 2014. Estimating standing biomass in papyrus (*Cyperus papyrus* L.)

swamp: Exploratory of in situ hyperspectral indices and random forest regression. *Int. J. Remote Sens.* 35, 693–714. <https://doi.org/10.1080/01431161.2013.870676>

Atera, E.A., Ishii, T., Onyango, J.C., Itoh, K., Azuma, T., 2013. Striga Infestation in Kenya: Status, Distribution and Management Options 2, 99–108. <https://doi.org/10.5539/sar.v2n2p99>

Baloloy, A., Blanco, A., Candido, C., Argamosa, R.J., Dumalag, J.B.L., Dimapilis, L.L., Paringit, E., 2018. Estimation of Mangrove Forest Aboveground Biomass Using Multispectral Bands, Vegetation Indices and Biophysical Variables Derived from Optical Satellite Imageries: RapidEye, PlanetScope and Sentinel-2. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* 4, 29–36. <https://doi.org/10.5194/isprs-annals-IV-3-29-2018>

Bannari, A., Morin, D., Bonn, F., Huete, A.R., 1995. A review of vegetation indices. *Remote Sens. Rev.* 13, 95–120. <https://doi.org/10.1080/02757259509532298>

Baugh, W.M., Groeneveld, D.P., 2006. Broadband vegetation index performance evaluated for a low-cover environment. *Int. J. Remote Sens.* 27, 4715–4730. <https://doi.org/10.1080/01431160600758543>

Breiman, L., 2002. Manual on setting up, using, and understanding random forests v3. 1. Tech. Report, <http://oz.berkeley.edu/users/breiman>, Stat. Dep. Univ. Calif. Berkeley, ... 29. <https://doi.org/10.2776/85168>

Breiman, L., 2001. Randomforest2001 1–33. <https://doi.org/10.1017/CBO9781107415324.004>

Cao, J., Liu, K., Zhu, Y., Li, J., He, Z., 2018. Identifying Mangrove Species Using Field Close-Range Snapshot Hyperspectral Imaging and Machine-Learning Techniques. *Remote Sens.* 10, 2047. <https://doi.org/10.3390/rs10122047>

Chemura, A., Mutanga, O., Dube, T., 2017. Integrating age in the detection and mapping of incongruous patches in coffee (*Coffea arabica*) plantations using multi-temporal Landsat 8 NDVI anomalies. *Int. J. Appl. Earth Obs. Geoinf.* 57, 1–13. <https://doi.org/10.1016/j.jag.2016.12.007>

Chemura, A., Mutanga, O., Odindi, J., Kutwayo, D., 2018. Mapping spatial variability of foliar nitrogen in coffee (*Coffea arabica* L.) plantations with multispectral Sentinel-2 MSI data. *ISPRS J. Photogramm. Remote Sens.* 138, 1–11. <https://doi.org/10.1016/j.isprsjprs.2018.02.004>

Cooley, S.W., Smith, L.C., Stepan, L., Mascaro, J., 2017. Tracking dynamic northern surface water changes with high-frequency planet CubeSat imagery. *Remote Sens.* 9, 1–21. <https://doi.org/10.3390/rs9121306>

Deng, H., 2013. Guided Random Forest in the RRF Package 1–3. <https://doi.org/10.1016/j.neuropsychologia.2011.12.015>

Deng, H., Runger, G., 2013. Gene selection with guided regularized random forest. *Pattern Recognit.* 46, 3483–3489. <https://doi.org/10.1016/j.patcog.2013.05.018>

Dube, T., Mutanga, O., Abdel-Rahman, E.M., Ismail, R., Slotow, R., 2015. Predicting Eucalyptus spp. stand volume in Zululand, South Africa: an analysis using a stochastic gradient boosting

- regression ensemble with multi-source data sets. *Int. J. Remote Sens.* 36, 3751–3772. <https://doi.org/10.1080/01431161.2015.1070316>
- Dube, T., Mutanga, O., Elhadi, A., Ismail, R., 2014. Intra-and-inter species biomass prediction in a plantation forest: Testing the utility of high spatial resolution spaceborne multispectral rapideye sensor and advanced machine learning algorithms. *Sensors (Switzerland)* 14, 15348–15370. <https://doi.org/10.3390/s140815348>
- Evangelista, P.H., Stohlgren, T.J., Morisette, J.T., Kumar, S., 2009. Mapping invasive tamarisk (Tamarix): A comparison of single-scene and time-series analyses of remotely sensed data. *Remote Sens.* 1, 519–533. <https://doi.org/10.3390/rs1030519>
- Hadjimitsis, D.G., Papadavid, G., Agapiou, A., Themistocleous, K., Hadjimitsis, M.G., Retalis, A., Michaelides, S., Chrysoulakis, N., Toullos, L., I. Clayton, C.R., 2010. Atmospheric correction for satellite remotely sensed data intended for agricultural applications: Impact on vegetation indices. *Nat. Hazards Earth Syst. Sci.* 10, 89–95. <https://doi.org/10.5194/nhess-10-89-2010>
- Han, H., Guo, X., Yu, H., 2017. Variable selection using Mean Decrease Accuracy and Mean Decrease Gini based on Random Forest. *Proc. IEEE Int. Conf. Softw. Eng. Serv. Sci. ICSESS* 219–224. <https://doi.org/10.1109/ICSESS.2016.7883053>
- Houborg, R., McCabe, M.F., 2018a. Daily retrieval of NDVI and LAI at 3 m resolution via the fusion of CubeSat, Landsat, and MODIS data. *Remote Sens.* 10. <https://doi.org/10.3390/rs10060890>
- Houborg, R., McCabe, M.F., 2018b. A Cubesat enabled Spatio-Temporal Enhancement Method (CESTEM) utilizing Planet, Landsat and MODIS data. *Remote Sens. Environ.* 209, 211–226. <https://doi.org/10.1016/j.rse.2018.02.067>
- Immitzer, M., Vuolo, F., Atzberger, C., 2016. First experience with Sentinel-2 data for crop and tree species classifications in central Europe. *Remote Sens.* 8. <https://doi.org/10.3390/rs8030166>
- Khan, Z.R., Hassanali, A., Overholt, W., Khamis, T.M., Hooper, A.M., Pickett, J.A., Wadhams, L.J., Woodcock, C.M., 2002. Control of witchweed *Striga hermonthica* by intercropping with *Desmodium* spp., and the mechanism defined as allelopathic. *J. Chem. Ecol.* 28, 1871–1885. <https://doi.org/10.1023/A:1020525521180>
- Kyalo, R., Abdel-Rahman, E.M., Subramanian, S., Nyasani, J.O., Thiel, M., Jozani, H., Borgemeister, C., Landmann, T., 2017. Maize Cropping Systems Mapping Using RapidEye Observations in Agro-Ecological Landscapes in Kenya. *Sensors* 17, 2537. <https://doi.org/10.3390/s17112537>
- Masocha, M., Dube, T., Mpofu, N.T., Chimunhu, S., 2018. Accuracy assessment of MODIS active fire products in southern African savannah woodlands. *Afr. J. Ecol.* 56, 563–571. <https://doi.org/10.1111/aje.12494>
- Matongera, T.N., Mutanga, O., Dube, T., Sibanda, M., 2017. Detection and mapping the spatial distribution of bracken fern weeds using the Landsat 8 OLI new generation sensor. *Int. J. Appl. Earth Obs. Geoinf.* 57, 93–103. <https://doi.org/10.1016/j.jag.2016.12.006>
- Midega, C.A.O., Wasonga, C.J., Hooper, A.M., Pickett, J.A., Khan, Z.R., 2017. Drought-tolerant *Desmodium* species effectively suppress parasitic striga weed and improve cereal grain yields in western Kenya. *Crop Prot.* 98, 94–101. <https://doi.org/10.1016/j.cropro.2017.03.018>
- Mureriwa, N., Adam, E., Sahu, A., Tesfamichael, S., 2016. Examining the spectral separability of *Prosopis glandulosa* from co-existent species using field spectral measurement and guided regularized random forest. *Remote Sens.* 8. <https://doi.org/10.3390/rs8020144>
- Mutanga, O., Dube, T., Galal, O., 2017. Remote Sensing of Crop Health for Food Security in Africa: Potentials and Constraints. *Remote Sens. Appl. Soc. Environ.* 8, 231–239. <https://doi.org/10.1016/j.rsase.2017.10.004>
- Planet, 2018. Planet — Planet Imagery Products.
- R Core Team, 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing.
- Ramoelo, A., Cho, M., Mathieu, R., Skidmore, A.K., 2015. Potential of Sentinel-2 spectral configuration to assess rangeland quality. *J. Appl. Remote Sens.* 9, 094096. <https://doi.org/10.1117/1.JRS.9.094096>
- Royimani, L., Mutanga, O., Odindi, J., Zolo, K.S., Sibanda, M., Dube, T., 2019. Distribution of *Parthenium hysterophoru* L. with variation in rainfall using multi-year SPOT data and random forest classification. *Remote Sens. Appl. Soc. Environ.* 13, 215–223. <https://doi.org/10.1016/j.rsase.2018.11.007>
- Scholes, J.D., Press, M.C., 2008. *Striga* infestation of cereal crops - an unsolved problem in resource limited agriculture. *Curr. Opin. Plant Biol.* 11, 180–186. <https://doi.org/10.1016/j.pbi.2008.02.004>
- Shoko, C., Mutanga, O., 2017. Examining the strength of the newly-launched Sentinel 2 MSI sensor in detecting and discriminating subtle differences between C3 and C4 grass species. *ISPRS J. Photogramm. Remote Sens.* 129, 32–40. <https://doi.org/10.1016/j.isprsjprs.2017.04.016>
- Sibanda, M., Mutanga, O., Rouget, M., 2015. Examining the potential of Sentinel-2 MSI spectral resolution in quantifying above ground biomass across different fertilizer treatments. *ISPRS J. Photogramm. Remote Sens.* 110, 55–65. <https://doi.org/10.1016/j.isprsjprs.2015.10.005>
- Thamaga, K.H., Dube, T., 2018. Remote sensing of invasive water hyacinth (*Eichhornia crassipes*): A review on applications and challenges. *Remote Sens. Appl. Soc. Environ.* 10, 36–46. <https://doi.org/10.1016/j.rsase.2018.02.005>
- Veloso, A., Mermoz, S., Bouvet, A., Le Toan, T., Planells, M., Dejoux, J.F., Ceschia, E., 2017. Understanding the temporal behavior of crops using Sentinel-1 and Sentinel-2-like data for agricultural applications. *Remote Sens. Environ.* 199, 415–426. <https://doi.org/10.1016/j.rse.2017.07.015>
- Vrieling, A., Meroni, M., Darvishzadeh, R., Skidmore, A.K., Wang, T., Zurita-Milla, R., Oosterbeek, K., O'Connor, B., Paganini, M., 2018. Vegetation phenology from Sentinel-2 and field cameras for a Dutch barrier island. *Remote Sens. Environ.* 215. <https://doi.org/10.1016/j.rse.2018.03.014>

Xue, J., Su, B., 2017. Significant remote sensing vegetation indices: A review of developments and applications. *J. Sensors* 2017. <https://doi.org/10.1155/2017/1353691>

Yue, S., Wenjiang, H., Huichun, Y., Chao, R., Naichen, X., Yun, G., Yingying, D., Dailiang, P., 2018. Partial Least Square Discriminant Analysis Based on Normalized Two-Stage Vegetation Indices for Mapping Damage from Rice Diseases Using 1–16. <https://doi.org/10.3390/s18061901>