Int J Appl Earth Obs Geoinformation 85 (2020) 102008

Contents lists available at ScienceDirect

Int J Appl Earth Obs Geoinformation

journal homepage: www.elsevier.com/locate/jag

Is it possible to discern Striga weed (*Striga hermonthica*) infestation levels in maize agro-ecological systems using *in-situ* spectroscopy?

Bester Tawona Mudereri[a,b,*], Timothy Dube[b], Saliou Niassy[a], Emily Kimathi[a], Tobias Landmann[a,d], Zeyaur Khan[a], Elfatih M. Abdel-Rahman[a,c]

[a] International Centre of Insect Physiology and Ecology (icipe), P.O. Box 30772, 00100, Nairobi, Kenya
[b] Department of Earth Sciences, University of Western Cape, Private Bag X17, Bellville, 7535, South Africa
[c] Department of Agronomy, Faculty of Agriculture, University of Khartoum, Khartoum North, 13314, Sudan
[d] Remote Sensing Solutions GmbH, Dingolfinger Str. 9, 81673, Munich, Germany

ARTICLE INFO

Keywords:
Invasive weeds detection
Maize
In-situ hyperspectral data
Machine learning
Resampled Sentinel-2

ABSTRACT

The invasion by Striga in most cereal crop fields in Africa has posed a significant threat to food security and has caused substantial socioeconomic losses. Hyperspectral remote sensing is an effective means to discriminate plant species, providing possibilities to track such weed invasions and improve precision agriculture. However, essential baseline information using remotely sensed data is missing, specifically for the Striga weed in Africa. In this study, we investigated the spectral uniqueness of Striga compared to other co-occurring maize crops and weeds. We used the *in-situ* FieldSpec® Handheld 2™ analytical spectral device (ASD), hyperspectral data and their respective narrow-band indices in the visible and near infrared (VNIR) region of the electromagnetic spectrum (EMS) and four machine learning discriminant algorithms (i.e. random forest: RF, linear discriminant analysis: LDA, gradient boosting: GB and support vector machines: SVM) to discriminate among different levels of Striga (*Striga hermonthica*) infestations in maize fields in western Kenya. We also tested the utility of Sentinel-2 waveband configurations to map and discriminate Striga infestation in heterogenous cereal crop fields. The *in-situ* hyperspectral reflectance data were resampled to the spectral waveband configurations of Sentinel-2 using published spectral response functions. We sampled and detected seven Striga infestation classes based on three flowering Striga classes (low, moderate and high) against two background endmembers (soil and a mixture of maize and other co-occurring weeds). A guided regularized random forest (GRRF) algorithm was used to select the most relevant hyperspectral wavebands and vegetation indices (VIs) as well as for the resampled Sentinel-2 multispectral wavebands for Striga infestation discrimination. The performance of the four discriminant algorithms was compared using classification accuracy assessment metrics. We were able to positively discriminate Striga from the two background endmembers i.e. soil and co-occurring vegetation (maize and co-occurring weeds) based on the few GRRF selected hyperspectral vegetation indices and the GRRF selected resampled Sentinel-2 multispectral bands. RF outperformed all the other discriminant methods and produced the highest overall accuracy of 91% and 85%, using the hyperspectral and resampled Sentinel-2 multispectral wavebands, respectively, across the four different discriminant models tested in this study. The class with the highest detection accuracy across all the four discriminant algorithms, was the "exclusively maize and other co-occurring weeds" (>70%). The GRRF reduced the dimensionality of the hyperspectral data and selected only 9 most relevant wavebands out of 750 wavebands, 6 VIs out of 15 and 6 out of 10 resampled Sentinel-2 multispectral wavebands for discriminating among the Striga and co-occurring classes. Resampled Sentinel-2 multispectral wavebands 3 (green) and 4 (red) were the most crucial for Striga detection. The use of the most relevant hyperspectral features (i.e. wavebands and VIs) significantly ($p \leq 0.05$) increased the overall classification accuracy and Kappa scores (±5% and ±0.2, respectively) in all the machine learning discriminant models. Our results show the potential of hyperspectral, resampled Sentinel-2 multispectral datasets and machine learning discriminant algorithms as a tool to accurately discern Striga in heterogenous maize agro-ecological systems.

* Corresponding author.
E-mail addresses: bmudereri@icipe.org (B.T. Mudereri), tidube@uwc.ac.za (T. Dube), sniassy@icipe.org (S. Niassy), ekimathi@icipe.org (E. Kimathi), Landmann@rssgmbh.de (T. Landmann), zkhan@icipe.org (Z. Khan), eabdel-Rahman@icipe.org (E.M. Abdel-Rahman).

https://doi.org/10.1016/j.jag.2019.102008
Received 31 May 2019; Received in revised form 22 October 2019; Accepted 1 November 2019
Available online 12 November 2019

1. Introduction

In Africa, food and nutrition insecurity due to crop losses is a chronic problem caused by insect pests, diseases, weeds and poor agronomic and soil management practices (Sasson, 2012). This food insecurity is likely to be worsened by the frequent unfavourable climatic conditions like droughts, climate change and variability, among others (Rakotoarisoa et al., 2012). The most important staple crops on the African continent that secure food and nutrient to about 1.2 billion people are maize, sorghum, wheat, millet, and rice (FAO et al., 2018). Among these economically important crops, maize plays the major role on the livelihood of people in sub-Saharan Africa (SSA). However, the productivity of maize has been on the decline in the last decade in SSA, mainly due to the emerging of invasive pests and diseases such as stemborers, fall armyworm, maize lethal necrosis and invasive weeds like Striga (FAO et al., 2018; Khan et al., 2014).

Striga, commonly referred to as the "witch weed" is considered to be the most economically important parasitic weed globally (Unachukwu et al., 2017). This parasitic weed attaches to the roots of the host plants after germination and causes considerable photosynthetic and productivity interference (Khan et al., 2002). Of the 23 Striga species predominant in Africa, *Striga hermonthica* is the most destructive, affecting a widespread range of crops including maize, sorghum, millet, rice, and sugarcane (Ejeta and Gressel, 2007). Striga can reduce cereal production as much as 20%–100% to more than 40 million households every year across Africa (Atera et al., 2013; Scholes and Press, 2008). Although these socioeconomic losses are difficult to quantify, it is estimated that in Africa alone, over US$ 1 billion is lost every year due to Striga infestation (Ejeta and Gressel, 2007; Spallek et al., 2013). Smallholder farmers are the most affected since they cannot afford the expensive Striga control mechanisms currently available on the market. These farmers often resort to the inefficient hand weeding aimed at reducing the Striga seed bank within the soil, which is unsustainable. This problem is aggravated by the viability of Striga seeds in the soil for up to 20 years and their complex potential to spread *via* both mechanical and cultural processes (Khan et al., 2002).

Due to the destructive nature of Striga, numerous technological and research developments have been made to help control or minimize Striga impacts on crop production. So far, efforts to control Striga have focused on the manipulation of genetics, chemical ecology and phenology of the weed (Midega et al., 2017; Oswald, 2005; Rispail et al., 2007; Samejima et al., 2016). However, on-farm Striga control technologies require spatiotemporal information on the weed to precisely prioritize sites for intervention and applications of such technologies. Usually, ground-based surveys and inspection methods are used to detect Striga-infested farms. This approach is often expensive, has a long-time lag, is laborious and provides incomplete information on Striga hotspots. In contrast, remote sensing provides efficient, timely, synoptic and inexpensive data that could effectively capture weeds spectral phenological responses at different spatiotemporal scales (Mutanga et al., 2017). Studies have shown that weeds distribution and abundance can be estimated using diverse types of sensors and instruments such as field based automated sensors (Smith and Blackshaw, 2003), unmanned aerial vehicles (de Castro et al., 2018; Peña et al., 2013), airborne multispectral and hyperspectral remote sensing (Mirik et al., 2013) among others. Yet, essential baseline information for the usage of such remote sensing information is absent for many high-impact invasive parasitic weeds like Striga (Große-Stoltenberg et al., 2016). In this study, we explored the potential of using *in-situ* hyperspectral remotely sensed data at plot level, to monitor Striga infestation in maize crops grown in an agro-ecological landscape in Kenya. We essentially tested whether canopy level *in-situ* hyperspectral data could discriminate among different Striga infestation levels and their co-occurring maize crop and other weeds. We further tested the potential capability of the Sentinel-2 multispectral band settings to detect and predict Striga infestation intensity, at plot level, in heterogenous cereal crop fields.

Hyperspectral instruments acquire data in numerous quasi-contiguous spectral wavebands, allowing detection of the spectral features of plant biochemical and physical characteristics like pigments, nutrients and water which are often masked when using the broadband multispectral data (Abdel-Rahman et al., 2013; Landmann et al., 2015). Hence, hyperspectral data are efficient in discriminating weed species from their co-occurring crops based on their biochemical and physical characteristics providing vast potential to precision farming for weed management (Große-Stoltenberg et al., 2016; Mureriwa et al., 2016). Additionally, *in-situ* hyperspectral data capture subtle spectral differences that are spectrally less distinct in airborne and spaceborne data (Sibanda et al., 2015). Thus, *in-situ* hyperspectral platforms enable quick spectral measurements of targets on the ground and offer the opportunity for band specific indices that breakdown complexes concealed in biochemical and physical characteristics of plants (Huang et al., 2015). These *in-situ* hyperspectral data, are also operated under chosen appropriate atmospheric conditions unlike when operating satellite sensors (Chen et al., 2009). This enables quality detection of unmixed energy captured from target objects without the influence of the bidirectional and diffuse scattering effects from other non-target features and the atmosphere (Jia et al., 2011). Also, the *in-situ* hyperspectral platforms can acquire spectral data at finer spatial resolution (up to a sub-metre), capturing the spectral vegetation signals at levels of a plant or an assemblage of plants. Such fine-scale remotely sensed data offer deeper understanding of the interaction between parasitic weeds like Striga and the electromagnetic radiation at ground level prior to upscaling to airborne or spaceborne platforms such as Sentinel-2 (Kumar et al., 2001).

The relatively new generation of multispectral spaceborne sensors such as Sentinel-2, have assumed the use of relatively narrower wavebands (*e.g.* 15 nm spectral width), including those in the red-edge region of the electromagnetic spectrum (EMS) centred at 705, 740 and 783 nm that were not present in previous broadband sensors like Landsat 7, 8 and the advanced spaceborne thermal emission and reflection (ASTER: Chemura et al., 2017). Therefore, there has been a growing interest to test the Sentinel-2 data, regarding its potential to advance precision agriculture and other operational uses, particularly in low income regions (Dhau et al., 2018; Mudereri et al., 2019). This is mainly because Sentinel-2 data are freely available, with a relatively higher spatial resolution (10 m) and possess strategically placed bands at the red-edge region of the EMS, which makes the sensor versatile for many applications (Ochungo et al., 2019). Therefore, citing these positive characteristics, Sentinel-2 is hypothesised to be capable of providing timely data for the generation of critical products for Striga monitoring.

Despite the previously mentioned advantages posed by both hyperspectral and Sentinel-2 multispectral datasets, these datasets alone might not be adequate for detecting Striga infestation in complex and heterogenous croplands. Merging the magnitude of the detail provided by hyperspectral data and the strength and capabilities of machine learning algorithms provides opportunities to reveal these complex structural and biophysical characteristics of weeds. However, one of the prominent problems in hyperspectral data processing and analysis is the dimensionality and multicollinearity inherent in the data (Adam et al., 2017). Multicollinearity is associated with the limited number of training samples (n) in contrast to the abundance of hyperspectral wavebands (p), that often hinder the performance of the predictive models when they are validated using independent test dataset (i.e. overfitting) (Adam et al., 2017; Mureriwa et al., 2016). Studies have utilized robust machine learning (ML) classification algorithms like support vector machines (SVM, Vapnik, 1979), linear discriminant analysis (LDA, Fisher, 1936), gradient boosting (GB, Friedman, 1999) and random forest (RF, Breiman, 2001) to deal with both the dimensionality and multicollinearity problems in the hyperspectral data. All these mentioned ML classifiers are assumption-free methods that do not encounter variable overfitting challenges and yield a variable

importance by-product which enables selection of fewer, yet relevant input predictors (i.e. Striga weed). Specifically, the guided regularized random forest (GRRF) and RF have shown to be successful methods in reducing the dimensionality of the hyperspectral data and simultaneously handle the multicollinearity in the data (Adam et al., 2017; Deng and Runger, 2013; Mureriwa et al., 2016). Nevertheless, previous studies demonstrated no consensus on the best ML classification algorithm and the best dimension reduction technique for invasive weeds discrimination (Große-Stoltenberg et al., 2016; Maxwell et al., 2018).

In this study, our innovation hinges on the hypothesis that flowering Striga is conspicuous from the rest of the photosynthetically green vegetation through their unique anthocyanins purple pigment in the flowers. To the best of our knowledge, the use of hyperspectral data and multivariate ML predictive models to separate different levels of Striga infestation using specifically the floral signal have not been examined. The present study was conducted with the following objectives:

a) To investigate the spectral uniqueness and behaviour of flowering and non-flowering Striga owing to varying levels of infestation and co-occurring vegetation (i.e. maize crop and other chlorophyll-active materials), and
b) To discriminate among different levels of Striga infestations in maize crop using the most relevant hyperspectral and resampled Sentinel-2 multispectral features and ML classification algorithms

2. Methods

2.1. Study site

The study was conducted in Rongo subcounty which lies within the Migori county in western Kenya. The study area is bound by the coordinates 0^0 39'12" S; 34^0 35'.40" E and 0^0 59'16" S; 34^0 37'21" E (Fig. 1) at an altitude of 1470 m above sea level. The climate in the study area is tropical and characterized by a yearly bimodal rainfall model with an average annual rainfall of 1600 mm across the two rainy seasons i.e. during "long rains" season occurring between March and June and a "short rains" season spanning November to January. The annual average temperature is 20.6^0 C and the relative humidity ranges between 50% and 70% while the soil type is loam, sandy and clay. The agro-natural ecosystem in the study area is dominated by scattered savanna grasslands in combination with deciduous and exotic forest vegetation, while the agricultural activities are mainly subsistence and small-scale farming. The crops grown in Rongo subcounty include sugarcane as the main cash crop, maize, bean, groundnut, green gram, cassava and some horticultural crops such as mango, banana, avocado, pawpaw and indigenous vegetables. Maize in Rongo subcounty area is grown as a mixed cropping system, with an average field size of 0.1 ha. The production of the crop in the study area is constrained mainly by rainfall variability and the invasive Striga weed. The yearly peak flowering period for Striga occurs between December and January during the short rains and again between May and June during the long rains season.

2.2. Field sampling design

In total, 70 quadrats were sampled within all representative fields for Striga infestations. We purposively sampled 14 fields during the period 12–16 December 2017 which coincided with peak Striga flowering window in the study area. Our purposive sampling procedure was guided by the presence and intensity of Striga infestation within each sampled maize field. In each sample maize field, we selected a plot of 30 m × 30 m and within each plot five quadrats measuring 1 m × 1 m each were laid out along two crossing diagonal transects. Specifically, two quadrats were laid out across each of the two diagonal transects and 10 m away from the plot edges while one quadrat was laid in the centre of the sample plot (Fig. 2).

In each quadrat, flowering and emerged Striga plants were counted. Infestation levels were categorized into three main classes; namely low (0–29 plants m^{-2}), moderate (30–90 plants m^{-2}), and high (>90 plants m^{-2}) Striga infestation classes. Specifically, our Striga infestation classes were characterized according to the average Striga population in each quadrat and the damage it causes to the maize crop; following the procedure described in Ekeleme et al. (2014). To test the influence of confounding features on the Striga spectral signal, we also collected spectral samples from soil background and a combination of maize and other weeds in the sample quadrat. This was necessary to test the influence of background spectral endmembers on the sensitivity of the Striga floral spectral signal. A total of seven classes of Striga infestation levels were assembled based on Striga floral signal sensitivity strength (number of Striga flowers per m^{-2}) and other EM abundances in the quadrat. These seven classes were derived from Striga infestation levels and corresponding combination background materials (soil or other non-Striga photosynthetically-active vegetation) in the sample quadrats. Firstly, we categorised three flowering Striga infestation classes (i.e. low: LW, moderate: MW and high: HW) with other green vegetation (maize and other weeds). The second category comprised pre-flowering (PF) Striga infestation in combination with green vegetation. The third category included exclusively the green vegetation (GV) class (maize and other weeds) with no Striga infestation. Again, we enumerated data for exclusively Striga within soil background (SB) with no green vegetation, and finally bare soil (SO). The spectra of Striga classes that occurred with a soil background (SB and SO) were collected by manually removing all non-Striga photosynthetically-active materials (for SB) and removing all vegetative material (for SO) from the sample quadrat. Table 1 summarises the Striga infestation class codes and their sample sizes. The variation in sample sizes was influenced by the availability of the different infestation levels within the selected sample quadrats.

2.3. In-situ hyperspectral data acquisition

Canopy-level *in-situ* hyperspectral data were collected within the sample quadrats using the FieldSpec® Handheld 2™ spectroradiometer (HH2: (ASD, 2010)) under clear skies and stable wind conditions at between 10:00 hrs and 14:00 hrs local time (Greenwich Mean Time: GTM + 3) as recommended by Sibanda et al. (2015). The field spectroradiometer used for the spectral data collection captures reflected radiation in 325–1075 nm of the EMS with a built-in 2 nm sampling resolution (ASD, 2010). The device then resamples the spectral data to 1 nm spectral range.

The hyperspectral measurements were collected from 1 m above the maize crop canopy using the bare optical input at nadir field of view. This covers an area of ~0.5 m in diameter on the target, which was enough for capturing the spectral signal of a group of maize and Striga plants. It is a rule of thumb that the diameter of a spot of light that is covered by the HH2 when it is perpendicularly positioned to a target is approximately half the distance of the instrument to any specific target area (FieldSpec et al., 2017). The instrument was held at arm's length (~0.9 m) from the observer to avoid scattered light from surrounding objects including the instrument and the operator (Kumar et al., 2013). The spectroradiometer was set to internally and automatically collect and average 20 spectral readings for each sample spectrum. In each of the 70 quadrats, we measured five spectra after optimizing and calibrating the measured radiance using a Spectralon white reference of ~100% reflectance. The optimization and calibration were done before a first measurement and after collecting the spectra of each sampling unit (i.e. quadrat), or when the instrument saturated because of changing ambient weather conditions like sun irradiance (FieldSpec et al., 2017). The final total of averaged spectra for each respective class that were used in this study are summarised in Table 1.

The spectral measurements acquired using the ASD were filtered using the "noiseFiltering" function and smoothened using the

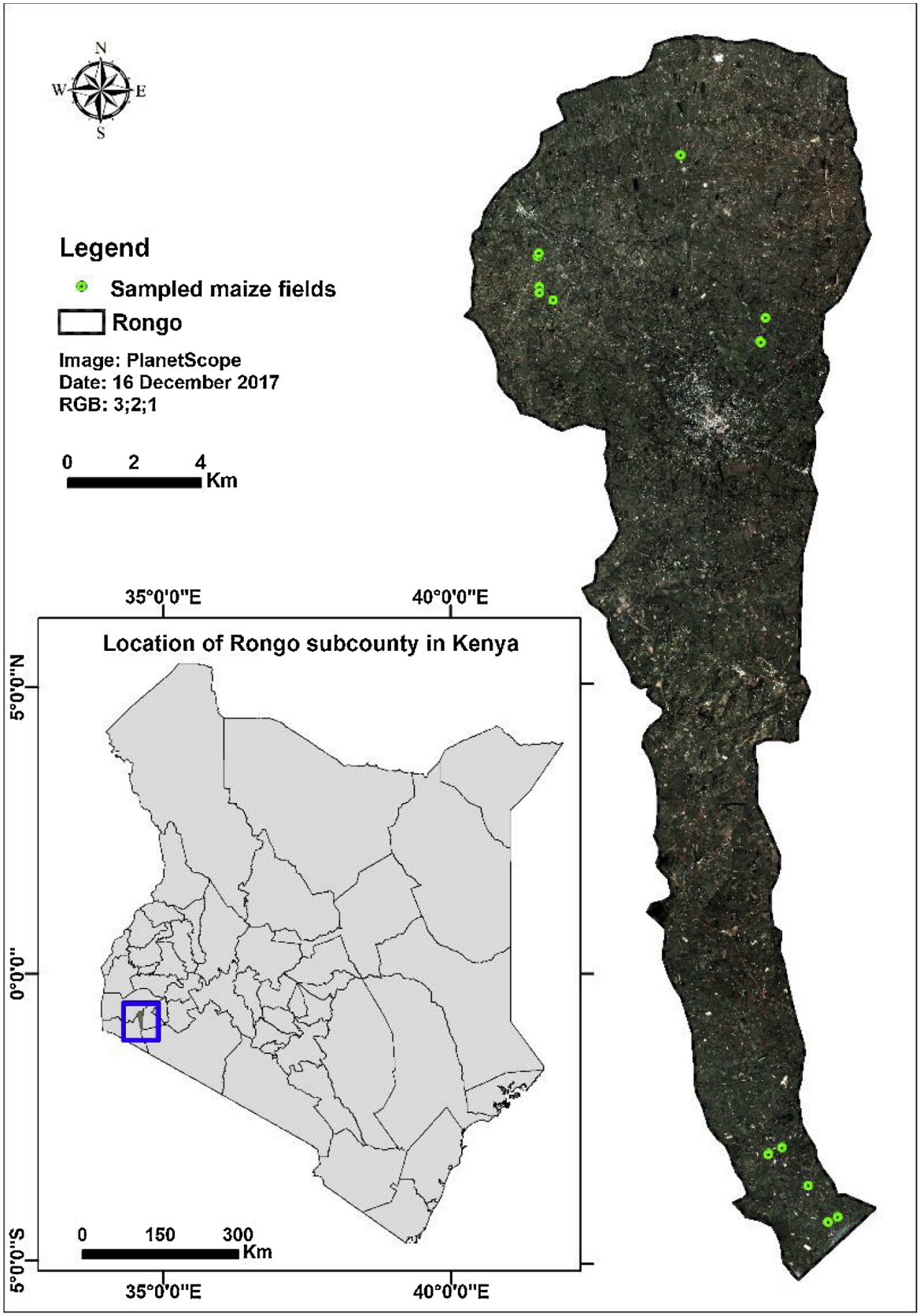

Fig. 1. Location of Rongo subcounty in Migori county, Kenya and the distribution of the sampled maize fields (n = 14). The image in the background is a PlanetScope image acquired on 16th of December 2017 and displayed in RGB: red (band3), green (band2) and blue (band1).

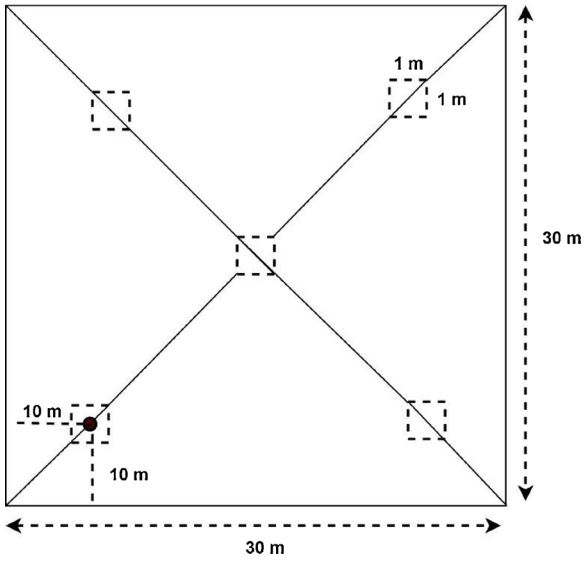

Fig. 2. Example structure of the distribution of quadrats within a 30 m × 30 m maize plot used for Striga sampling data collection.

"Savitzky–Golay" filter in the "hsdar" package (Lukas et al., 2018) in R software (R Core Team, 2018). These filtered spectra were resampled to the spectral configuration of Sentinel-2 using the spectral response function, i.e. "SpectralResampling" of Sentinel-2 present in the "hsdar" package. The Sentinel-2 multispectral wavebands description, waveband centres and their respective spectral wave ranges are shown in Table 2.

Fig. 3 shows the average spectral responses of each of the seven studied classes within the raw spectra 325–1075 nm range (i.e. 750 wavebands) of the EMS and the reflectance of the resampled Sentinel-2 spectra. Spectra at each of the wavebands were utilized as predictor variables to discriminate among the seven Striga infestation classes.

2.4. Calculation of the narrow-waveband vegetation indices

In addition to the 750 wavebands, we also calculated fifteen narrow-waveband vegetation indices (VIs) (Table 3) using the "hsdar" package (Lukas et al., 2018) in R software (R Core Team, 2018) and used them as predictor variables to discriminate among the seven Striga infestation classes. These indices were selected based on their relatedness to specific plant biophysical parameters (*e.g.* floral signal strength, plant health condition, plant pigments and plant water content) and the availability of the narrow-wavebands used in their formulae in our hyperspectral data that ranged between 325 nm and 1075 nm.

2.5. Predictor variables' selection using guided regularized random forest (GRRF) algorithm

We used GRRF to select fewer, yet the most relevant narrow-wavebands, VIs and resampled Sentinel-2 multispectral wavebands to discriminate the seven Striga infestation classes. We used the package "RRF" in R software (Deng, 2013; R Core Team, 2018). The regularized framework considerably reduces the training time by building a single model (Deng and Runger, 2013). The GRRF uses the same concept of a RF model but uses the importance scores generated from RF to guide the variable selection process (Mudereri et al., 2019; Mureriwa et al., 2016). The importance score of a variable in RF is obtained through the "Gini index" over all nodes across all RF decision trees obtained, and the variable is used to measure the purity of the feature at every node to facilitate the voting process of RF trees (Breiman, 2002). Compared to the variable importance feature in the ordinary RF, GRRF provides the precise variables that are most suitable for predicting the feature from the multiple features data set (Deng, 2013). GRRF uses a gamma value to penalise the selection of new features over features already selected that possess similar gain (importance). The gamma value occurs between 0 and 1 with values closer to 1 executing higher penalties, hence selecting fewer relevant variables within GRRF. Comparatively, the values closer and equal to 0 increase the number of potential relevant features selected, while the value of 0 yields similar variables to those produced when using an ordinary regularised random forest (RRF) (Deng and Runger, 2013). In this study, we used a *gamma* (γ) value of 0.8 to limit the variables (i.e. narrow-wavebands or VIs) selection. Our choice of $\gamma = 0.8$ was conservative as the highest gamma value of 1, extremely reduced the variables to too few (n = 3). The raw importance scores obtained from RF are normalized for each feature using Eqs. (1) to (3) to get the score used for variable selection in GRRF. For a detailed explanation of the theoretical and mathematical background of GRRF and how it functions, the readers are referred to Deng and Runger (2013).

$$\text{gain}_G\,(Xi) = \lambda i\ \text{gain}\,(Xi) \quad (1)$$

Table 2
The wavebands, waveband centres and their respective spectral width of the Sentinel-2 multispectral sensor. The wavebands that correspond to the *in-situ* hyperspectral data used in this study are shown in bold.

Waveband	Waveband description	Waveband centre (nm)	Wave range (nm)
1	**Coastal aerosol**	**443**	**433–453**
2	**Blue**	**490**	**458–523**
3	**Green**	**560**	**543–578**
4	**Red**	**665**	**650–680**
5	**Red-edge 1**	**705**	**698–713**
6	**Red-edge 2**	**740**	**733–748**
7	**Red-edge 3**	**783**	**773–793**
8	**Near Infrared (NIR)**	**842**	**785–900**
8a	**Near Infrared narrow (NIRn)**	**865**	**855–875**
9	**Water vapour**	**945**	**935–955**
10	Shortwave Infrared (cirrus)	1380	1360–1390
11	Shortwave Infrared 1 (SWIR1)	1610	1565–1655
12	Shortwave Infrared 2 (SWIR2)	2190	2100–2280

Table 1
Striga infestation level classes, with their respective class descriptions, class codes, sample sizes, training and testing samples used for employing the classification machine learning algorithms.

Class description	Class code	Sampled spectra	Train	Test
Maize and other weeds (green vegetation) with no Striga infestation	GV	32	22	10
High Striga infestation level with other green vegetation	HW	101	70	31
Moderate Striga infestation level with other green vegetation	MW	71	50	21
Low Striga infestation level with other green vegetation	LW	56	40	16
Pre-flowering Striga with other green vegetation	PF	20	15	5
Exclusive Striga within a soil background with no other green vegetation	SB	21	15	6
Bare soil devoid of any photosynthetic material	SO	20	15	5

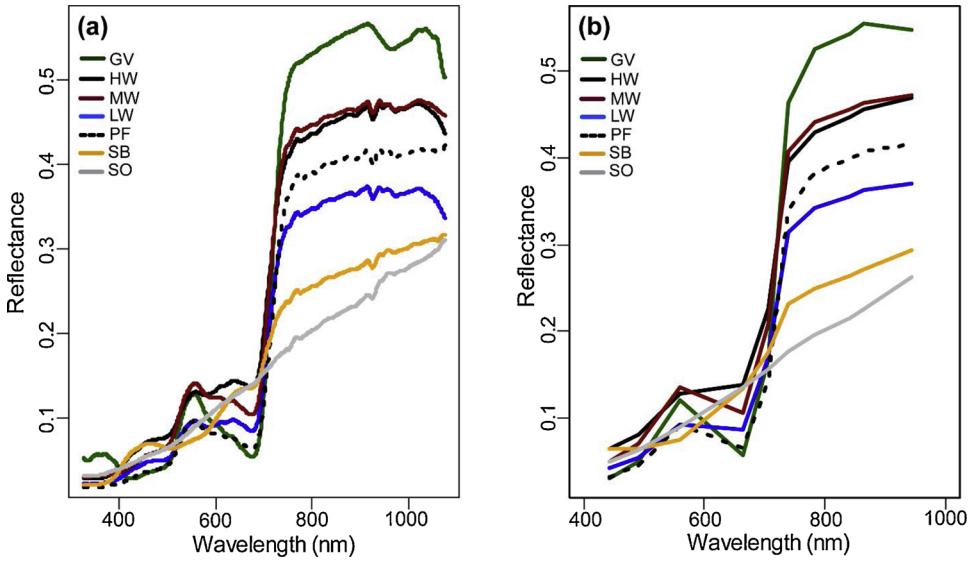

Fig. 3. Mean canopy-level spectra of (a) *in-situ* hyperspectral reflectance and (b) resampled Sentinel-2 multispectral reflectance of the seven studied classes: maize and other weeds (green vegetation) with no Striga infestation (GV); high Striga infestation level with other green vegetation (HW); moderate Striga infestation level with other green vegetation (MW); low Striga infestation level with other green vegetation (LW); pre-flowering Striga with other green vegetation (PF); exclusive Striga stands within a soil background with no other green vegetation (SB); bare soil devoid of any photosynthetic material (SO) measured using Hand Held FieldSpec® 2 (HH2) spectroradiometer in the 325–1075 nm wave range of the electromagnetic spectrum.

where: *gain (Xi)* denote the Gini information gain of using a feature *Xi* to split a tree node. And λi is calculated as:

$$\lambda i = 1 - \gamma + \gamma\ (Impi\ /\ Imp\star) \quad (2)$$

where: *Impi* is the importance score of *Xi* from RF and *Imp** is the maximum importance score, Therefore:

$$gain_G\ (Xi) = (Impi\ /\ Imp\star)\ X\ gain\ (Xi) \quad (3)$$

2.6. Machine learning discriminant algorithms

Four ML discriminant models namely; GB, LDA, RF and SVM were used to discriminate the seven Striga infestation classes. We compared the performance of these four models in discriminating the seven Striga infestation classes using the *in-situ* hyperspectral data under the following five predictor variable criteria: (1) the clipped range (400–1075) of the collected narrow-wavebands (n = 676); (2) all the calculated narrow-waveband VIs shown in Table 3 (n = 15); (3) The GRRF selected narrow-wavebands (n = 9); (4) The GRFF selected narrow-waveband VIs (n = 6), and (5) a combination of both GRRF selected narrow-wavebands and narrow-waveband VIs (n = 15). Similarly, using the mentioned four ML discriminant models, we further compared the performance of (a) all the resampled Sentinel-2 multispectral wavebands (n = 10) and (b) the GRRF selected resampled Sentinel-2 multispectral wavebands (n = 6).

GB, LDA, RF and SVM algorithms were selected in this study, because they have been widely used in classifying vegetation-related classes with reasonably high classification accuracies when hyperspectral data sets were utilized (Abdel-Rahman et al., 2014; Dube and Mutanga, 2015; Ramoelo et al., 2015). Further, these four classification algorithms do not require the traditional regression assumptions which makes them useful in many cases (Holloway and Mengersen, 2018). They possess advantages such as: (1) easily identifying and adapting to inherent patterns and trends in data, (2) little to no human intervention in the running process, (3) versatile in handling ad-hoc multi-dimensional and multivariate types of data and (4) mapping classes with

Table 3
Hyperspectral narrow-waveband vegetation indices used in this study.

Vegetation index	Related to:	**Equation	Reference
Fluorescence ratio Blue/Red (SR7)	Fluorescence	R_{440}/R_{690}	(Große-Stoltenberg et al., 2016)
Water band index (WBI)	Water	R_{900}/R_{970}	(Ho, 2009)
Simple ratio pigment index (SRPI)	Pigments	R_{430}/R_{680}	(Große-Stoltenberg et al., 2016)
Double peak index (DPI)	Vegetation stress	$(R_{688} \times R_{710})/R^2_{697}$	(Große-Stoltenberg et al., 2016)
Anthocyanin reflectance index (ARI)	Anthocyanin	$(1/R_{500}) - (1/R_{700})$	(Ho, 2009)
Anthocyanin reflectance index 2 (ARI2)	Anthocyanin	$(1/R_{550}) - (1/R_{700})$	(Ho, 2009)
Datt4	Pigments	$R_{672}/ (R_{550} \times R_{708})$	(Große-Stoltenberg et al., 2016)
Plant Senescing reflectance index (PSRI)	Leaf senescence	$(R_{678} - R_{500}) / R_{750}$	(Große-Stoltenberg et al., 2016)
Double difference index (DDN)	Chlorophyll	$2 \times (R_{710} - R_{660} - R_{760})$	(Große-Stoltenberg et al., 2016)
Modified Simple ratio (mSR)	Chlorophyll	$(R_{800} - R_{445})/ (R_{680} - R_{445})$	(Sims and Gamon, 2002)
Structure insensitive pigment index (SIPI)	Pigments	$(R_{800} - R_{445})/ (R_{800} - R_{680})$	(Ho, 2009)
Photochemical reflectance index (PRI)	Carotenoid	$(R_{531} - R_{570})/ (R_{531} + R_{570})$	(Sims and Gamon, 2002)
Photochemical Reflection Index × Chlorophyll content (PRI.CI2)	Carotenoid	$(R_{531} - R_{570})/ (R_{531} + R_{570}) \times (R_{760}/R_{700} - 1)$	(Große-Stoltenberg et al., 2016)
Transformed Chlorophyll Absorption Ratio Index (TCARI2)	Chlorophyll	$3 \times ((R_{750} - R_{705}) - 0.2\ (R_{750} - R_{550})\ (R_{750}/R_{705})$	(Große-Stoltenberg et al., 2016)
Enhanced Vegetation Index (EVI)	Biomass/LAI	$2.5 \times ((R_{800} - R_{670})/ (R_{800} - (6 \times R_{670}) - (7.5 \times R_{475}) + 1)$	(Ho, 2009)

** R is reflectance at the respective hyperspectral narrow-waveband and LAI is the leaf area index.

Table 4
R software packages used by "Caret" that were used in this study and their respective caret syntax code.

Algorithm	Caret code	Package	Reference
Random forest	"rf"	Ranger	(Liaw et al., 2002)
Support vector machines	"svmRadial"	Kernlab	(Karatzoglou et al., 2004)
Linear discriminant analysis	"lda"	Mass	(Venables and Ripley, 2002)
Stochastic gradient boosting	"gbm"	gbm and plyr	(Greenwell et al., 2019)

complex characteristics (Maxwell et al., 2018).

The "Caret" package (Kuhn et al., 2018) in R software was used to run and validate all the four ML discriminant models. The "Caret" package provides a standard syntax to execute a variety of ML discriminant approaches, thus simplifying the procedure of systematically comparing different algorithms and approaches (Maxwell et al., 2018). For consistency the tune length parameter was set to 10, so that 10 values for each parameter were assessed. Also, all the variables were centred and rescaled for consistency prior to classification. Table 4 summarises the "Caret" packages used to execute the four algorithms: RF, SVM, LDA and GB.

Again, for consistency purposes, the comparison of the performance of the four ML discriminant algorithms, were evaluated using the same dataset for all the models, split into a training set (70%) for model training and a test set (30%) for validation of the models (Dube and Mutanga, 2015; Qiao et al., 2017). Model performance was presented using boxplots of overall accuracy and Kappa within the five different predictor variable criteria mentioned for the hyperspectral data and two variable grouping criteria of the resampled Sentinel-2 multispectral wavebands. The inter-class prediction performances of the different models were further assessed using confusion matrices derived from the best performance predictor variables category for each algorithm. The McNemar test for paired categorical data represented in contingency tables was performed at 95% confidence interval (CI) to compare the performance among the four models in their ability to predict Striga severity classes using the GRRF selected variables.

3. Results

3.1. Spectral behaviour of flowering and non-flowering Striga

Fig. 4 demonstrates the variation in spectral responses of the different classes according to flowering compaction and colour. The near-infrared (NIR: 750–1075 nm) revealed multiple scattering within the leaf structure, emanating from the different compositions of the classes. Nonetheless, the NIR displayed significant differences in magnitude with green vegetation producing a plateau of high reflectance compared to all other classes where pigments no longer absorb the radiation. The patterns of Striga infestation levels low, moderate and high in the whole spectrum were similar but differed in the magnitude (>5% difference between high and low). This is emphasized in Fig. 4(b) in which the pattern for the classes is the same, but the magnitude of the reflectance differs as influenced by the Striga floral compaction and colour. The higher the number of flowers in a plot, the more the reflectance magnitude increased within the region 500–700 nm. However, when compared to the other non-Striga classes, much variation was observed in the red section (620–680 nm). The hyperspectral reflectance values for the "high", "moderate" and "low" Striga infestation, all peak around 550 nm and 670 nm which corresponded to the resampled Sentinel-2 wavebands 2 and 4, respectively. However, there is a slight depression in the reflectance values around 680 nm proceeding to increase again in the red-edge and NIR waveband regions. When Striga infested plots were compared to GV, they all have peaks at 550 nm but there is a marked difference at 680 nm where there is a huge depression for the GV class. Similarly, within the blue region (400–500 nm), Striga infestation classes show higher values compared to the GV class. Furthermore, a similar trend is observed between flowering Striga and PF class (Fig. 4c). Although visually the red-edge (680–750 nm) does not show any considerable variations in the spectra, green vegetation had the steepest gradient. The red-edge remains crucial in calculating vegetation indices as the gradient of the graphs reveals biochemical and ecophysiological vegetation parameters.

3.2. Predictor variables selection

The GRRF was able to determine only 6 narrow-waveband VIs, 9 narrow-wavebands and 6 resampled Sentinel-2 multispectral wavebands to be of utmost relevance for discriminating the seven Striga infestation classes (Fig. 5). The VIs that were selected by GRRF as the most relevant predictor variables were mainly related to pigments (Datt4, ARI, ARI2 and PRI.CI2). The most important narrow-wavebands are well distributed across the VNIR electromagnetic spectrum. Many of the most relevant narrow-wavebands and resampled Sentinel-2 multispectral wavebands for discriminating among the infestation levels occurred within the green (band 3) and the red (band 4) regions of the EMS. Although most of these GRRF selected narrow-wavebands occurred within the visible range, the one with the highest variable importance value among them was identified within the red-edge region of the EMS (at 677 nm).

The GRRF algorithm was able to determine uncorrelated variables for the VIs, however most of the narrow-wavebands selected were correlated (Fig. 6). The most correlated (>80%) among the narrow-wavebands were those in the green region of the EMS, while Datt4 and DDN were also negatively correlated to most of the green, red and NIR wavebands of the EMS.

3.3. Striga infestation levels discrimination using the four machine learning GB, LDA, RF and SVM models

The results showed that RF algorithm outperformed (overall accuracy of 91% and Kappa of 0.84, Fig. 7) all the other three ML discriminant algorithms in discriminating among the seven Striga infestation classes using the hyperspectral data; and the resampled Sentinel-2 multispectral wavebands (overall accuracy 85% and Kappa of 0.80: Fig. 8). This performance was followed by GB, LDA, and SVM, respectively. The use of only the selected VIs for the hyperspectral data and the GRRF selected wavebands for the resampled Sentinel-2, resulted in more accurate Striga infestation discrimination results compared with the use of other predictor variables across the ML discriminant algorithms, except for LDA (Figs. 7 and 8). Although RF results showed superiority over all the other algorithms, the overlaps of the boxplots tests showed that there was no significant difference ($p \geq 0.05$) with the performance of the GB algorithm. The LDA and SVM algorithms achieved the least accurate Striga discrimination results to predict Striga infestation, using all the predictor variable criteria. Also, the Kappa statistic revealed that LDA and SVM models performed not significantly different ($p \geq 0.05$) from model performance at random, whereas both the RF and the GB models produced high Kappa statistic values (Kappa > 0.75) for discriminating the seven classes.

Considering the inter-class prediction accuracies (i.e. individual producer's accuracy (PA) and user's accuracy (UA) metrics), RF model

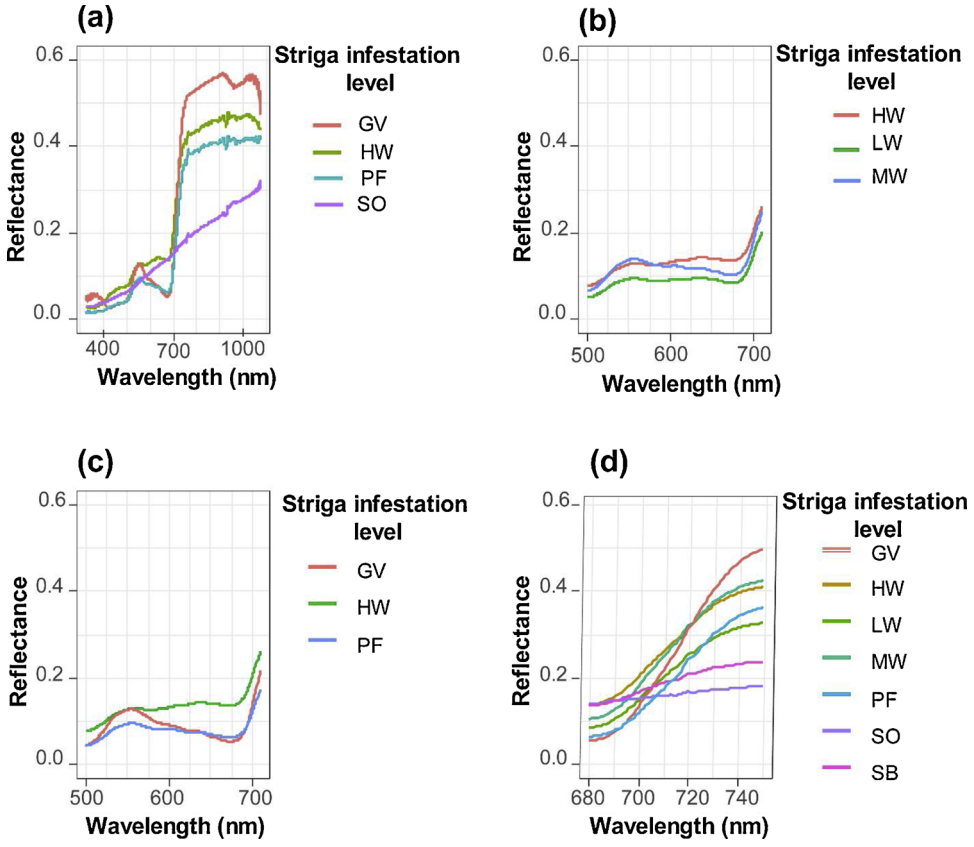

Fig. 4. Comparison of spectral behaviour for: (a) bare soil (SO), green vegetation (GV), high Striga infestation (HW) and pre-flowering (PF) Striga within the full spectral range (325–1075 nm); (b) high (HW), moderate (MW) and low (LW) Striga infestation levels within the visible range (500–700 nm); (c) green vegetation (GV), high Striga infestation (HW) and pre-flowering (PF) Striga within the visible range (500–700 nm); (d) all the seven classes used in this study within the red-edge spectral range (680–750 nm).

was superior over all the other models in predicting each of the seven classes using both the hyperspectral data or the resampled Sentinel-2 multispectral data (Tables 5 and 6). Although all the four ML discriminant models were able to predict HW with a PA of at least 70%, MW and LW were not consistent among the models. On the contrary, all the ML discriminant models were relatively reliable in predicting GV, SO and SB. Generally, the PA and UA metrics for LDA and SVM were relatively poor compared to the nonlinear decision tree-based algorithms (RF and GB) as shown in Table 5 and 6.

3.4. Pairwise model performance comparison using McNemar test

The performances of the four ML discriminant models in predicting Striga infestation were significantly different ($p \leq 0.05$) from each other, except the comparison between LDA and SVM (hyperspectral wavebands) and GB and SVM (resampled Sentinel-2 multispectral wavebands) when using the pairwise McNemar test (Table 7). This further confirms the superiority of RF in the prediction of Striga infestation when compared to GB, LDA and SVM.

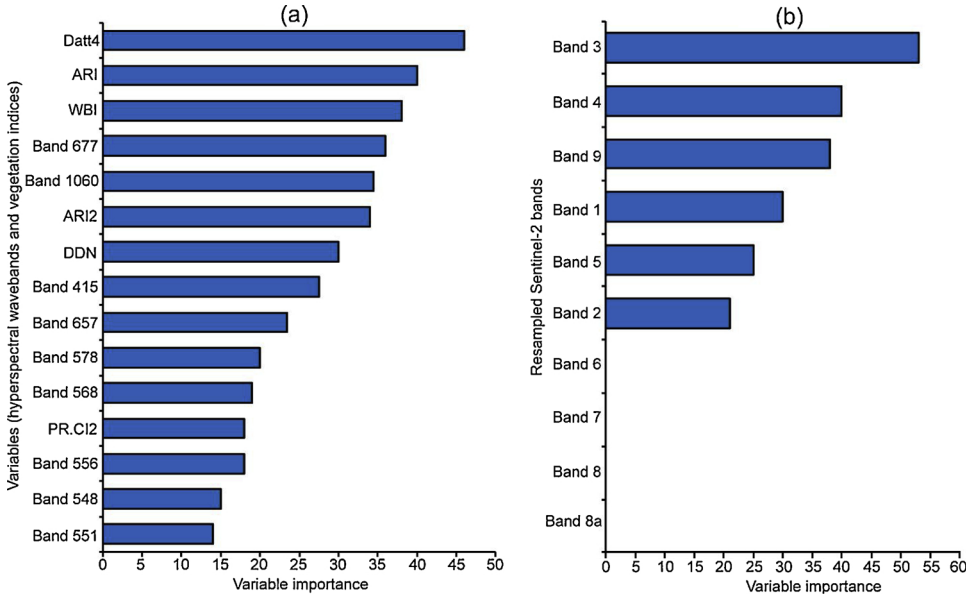

Fig. 5. Predictor variables relevance for (a) both hyperspectral narrow-waveband vegetation indices and narrow-wavebands selected using the variable selection measure of the guided regularized random forest (GRRF) algorithm (b) Resampled Sentinel-2 multispectral wavebands selected using the variable selection measure of the GRRF algorithm. See Table 2 for the descriptions of Sentinel-2 multispectral wavebands.

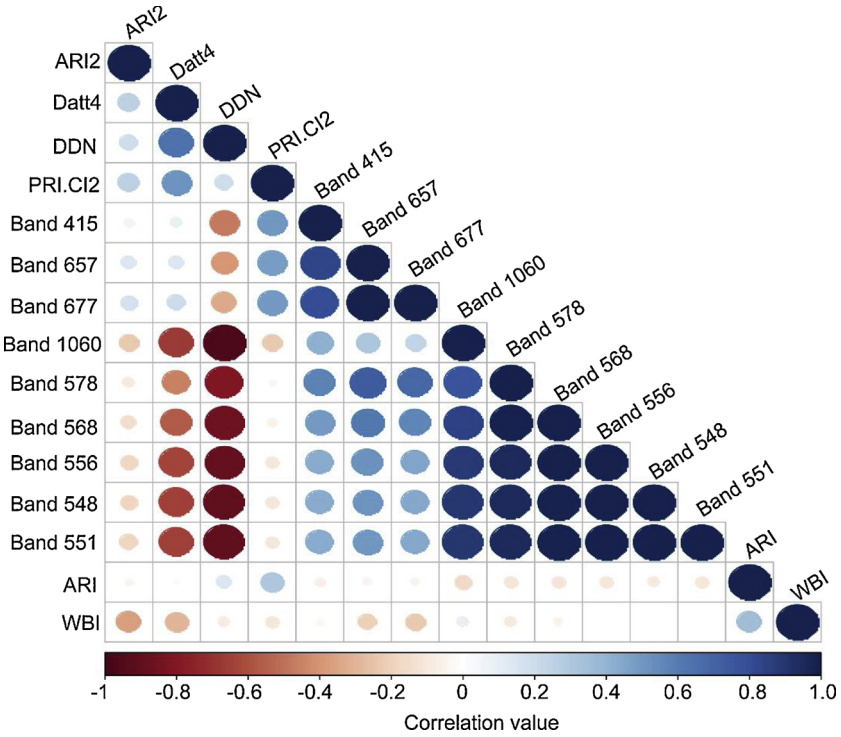

Fig. 6. Correlation matrix for the guided regularized random forest (GRRF) selected hyperspectral wavebands and indices. Darker shades of blue and red colors indicate high variable correlation, while light shades indicate low correlation between variables.

4. Discussion

Several studies have demonstrated the importance of managing and controlling Striga infestation and spatial spread (Atera et al., 2013; Khan et al., 2007; Spallek et al., 2013). This study examined the potential to use canopy-level *in-situ* hyperspectral data and resampled Sentinel-2 multispectral wavebands in distinguishing Striga from other co-occurring vegetated and non-vegetated materials within maize fields and to differentiate among different Striga infestation levels. Results from this current study show that Striga can accurately be discriminated from other vegetation and soil classes in maize fields using hyperspectral wavebands, narrow-band indices, resampled Sentinel-2 multispetral wavebands and machine learning discrimination algorithms.

4.1. Spectral behaviour of flowering and non-flowering Striga infestation classes and their co-occurring vegetation and soil

In this study, the results showed that the spectral behaviour of flowering Striga, non-flowering Striga and other green vegetation differ. This can be attributed to the diversity of plant pigments occurring at different levels within the different co-occurring flora. Plant pigments are inherently associated with the biological function of leaves. Chlorophylls absorb light energy and allocate it to the photosynthetic system while, yellow pigments (carotenoids) also contribute energy to the photosynthetic apparatus and assist in resistance to environmental stress (Blackburn, 2007). Additionally, anthocyanins (red, pink and purple pigments) may also serve as scavengers of reactive

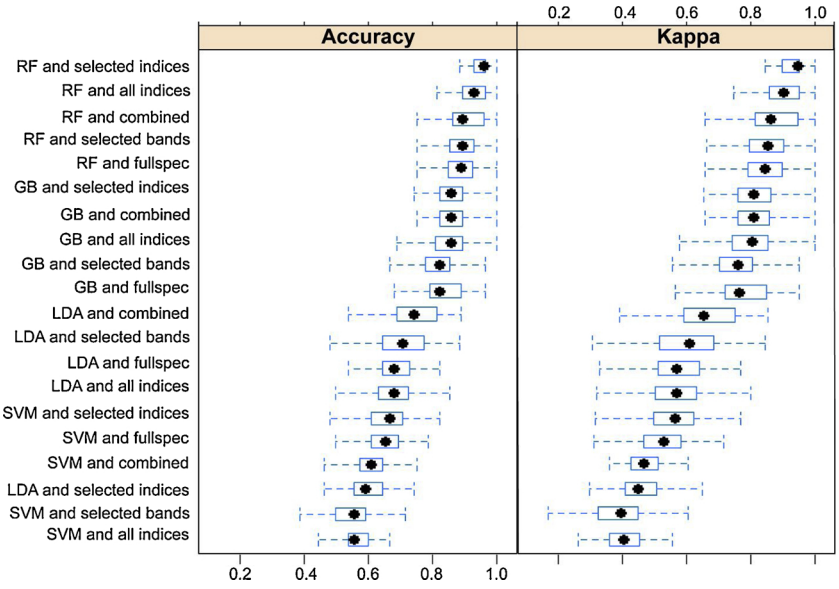

Fig. 7. Striga infestation discrimination models performance as evaluated by overall accuracy and Kappa statistics using the clipped range (400–1075 nm) of spectral narrow-wavebands (*fullspec), narrow-waveband vegetation indices (*indices), all indices and all narrow-wavebands (*combined), selected narrow-waveband indices (*selected indices) and selected narrow-wavebands (*selected bands). RF, GB, LDA and SVM are random forest, stochastic gradient boosting, linear discriminant analysis and support vector machines, respectively.

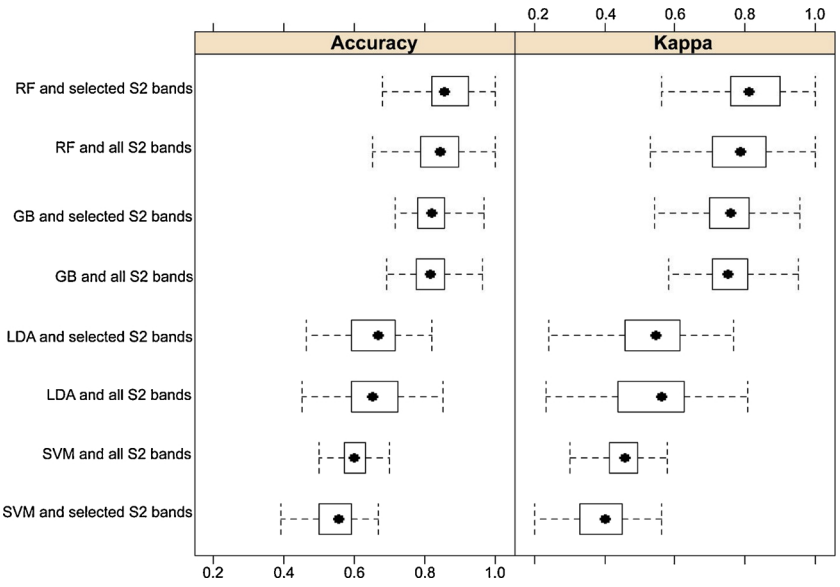

Fig. 8. Striga infestation discrimination models performance as evaluated by overall accuracy and Kappa statistics using all the resampled Sentinel-2 multispectral wavebands or only the 6 guided regularized random forest (GRRF) selected resampled Sentinel-2 multispectral wavebands. RF, GB, LDA and SVM are random forest, stochastic gradient boosting, linear discriminant analysis and support vector machines, respectively.

Table 5
Summarized confusion matrices and classification accuracies, overall accuracy (OA), producer's accuracy (PA) and user's accuracy (UA) of the random forest (RF), stochastic gradient boosting (GB), linear discriminant analysis (LDA) and support vector machines (SVM) discriminant models using the guided regularized random forest (GRRF) selected narrow-band indices.

	Machine learning algorithm							
	RF		GB		SVM		LDA	
Class: Infestation level	PA (%)	UA (%)	PA (%)	UA (%)	PA (%)	UA (%)	PA (%)	UA (%)
High (HW)	94	92	93	78	82	100	72	58
Moderate (MW)	85	92	78	84	100	10	41	44
Low (LW)	88	83	74	78	36	83	44	44
Green vegetation (GV)	100	100	70	100	100	100	100	100
Soil (SO)	100	100	100	100	0	0	100	100
Soil Background (SB)	100	100	86	100	0	0	71	83
Pre-flowering (PF)	100	100	75	100	33	33	30	67
OA (%)	91		83		60		55	

Table 6
Summarized confusion matrices and classification accuracies, overall accuracy (OA), producer's accuracy (PA) and user's accuracy (UA) of the random forest (RF), stochastic gradient boosting (GB), linear discriminant analysis (LDA) and support vector machines (SVM) discriminant models using the guided regularized random forest (GRRF) selected resampled Sentinel-2 multispectral wavebands.

	Machine learning algorithm							
	RF		GB		SVM		LDA	
Class: Infestation level	PA (%)	UA (%)	PA (%)	UA (%)	PA (%)	UA (%)	PA (%)	UA (%)
High (HW)	89	94	90	72	75	90	69	86
Moderate (MW)	87	76	70	85	50	10	76	37
Low (LW)	79	80	69	71	46	89	69	71
Green vegetation (GV)	92	100	80	90	50	100	40	33
Soil (SO)	100	100	100	100	100	100	100	100
Soil Background (SB)	95	100	94	100	100	30	83	100
Pre-flowering (PF)	75	75	78	80	0	0	40	67
OA (%)	85		81		65		55	

oxygen intermediates or as antifungal compounds (Litchenthaler and Buschmann, 2001; Sims and Gamon, 2002). In light of the importance of pigments for leaf and petal function, dynamics in pigment quantities may provide details regarding their physiological state (Thenkabail et al., 2013). For instance, chlorophylls tends to decline more rapidly than carotenoids when plants are under stress or during leaf senescence (Sims and Gamon, 2002). Similarly, the reflectance response to incident radiation is influenced by the quantity and the interplay between the ratios of these pigments (Blackburn, 2007). These could have led to the differences observed among our flowering Striga, non-flowering Striga and other vegetation classes.

We found that although the 530–570 nm are the portions which are mainly inclined to the green reflectance peak, all the plants had the highest peak at 550 nm within the visible region of the EMS, however, they differed significantly in the range 550–680 nm. This information is thus masked out, when using spaceborne sensors such as Sentinel-2 that

Table 7
McNemar test for comparing the performance of the four machine learning discriminant models in predicting the seven studied Striga infestation classes using the hyperspectral wavebands and the resampled Sentinel-2 multispectral wavebands. RF, GB, LDA and SVM are random forest, stochastic gradient boosting, linear discriminant analysis and support vector machines models, respectively.

	Hyperspectral wavebands		Resampled Sentinel-2 multispectral wavebands	
Comparison	Chi-square	*p*-value	Chi-square	*p*-value
RF vs GB	4.93	0.02*	5.14	0.02*
RF vs LDA	31.03	<0.001**	15.75	<0.001**
RF vs SVM	22.42	<0.001**	5.79	0.02*
GB vs LDA	19.86	<0.001**	7.84	<0.001**
GB vs SVM	10.26	0.001**	0.36	0.55
LDA vs SVM	0.21	0.64	6.05	0.01*

* Significant at 95% confidence interval (CI); ** Significant at 99% CI.

group these multiple narrow-wavebands into single broadbands i.e. in this case, into wavebands 3 and 4. Using the hyperspectral data revealed that the reflectance of our Striga classes within the 550–680 nm region of the EMS remained high, whilst the other green vegetation classes reflectance dropped considerably. This could be attributed to the presence and concentrations of anthocyanins within the petals of the Striga plants when flowering, because chlorophylls and carotenoids absorbance increases whereas reflectance by anthocyanins increases significantly in this regions of the EMS (Huang et al., 2015; Sims and Gamon, 2002). This concurs with the results of Blackburn (2007) who observed that, an increase in reflectance in the red region of the EMS was linked with an increase in anthocyanins. In addition, in the analysis among the three Striga infestation levels (HW, MW and LW), the magnitude of this reflectance increased with increase in flower compaction (number of flowering Striga m^{-2}) which confirmed the importance of this region for discriminating Striga from other photosynthetic-active plants. Leveraging on the presence of the purple colour pigment in the flowers proved crucial in the separation of Striga occurring with other similarly co-occurring weeds and crops. *S. hermonthica* flowers are purple (Ejeta and Gressel, 2007; Khan et al., 2002; Spallek et al., 2013), hence the violet section of the EMS is very decisive to distinguish flowering Striga from green plants' material. However, the downside to the violet section is the short wavelength characteristic which is easily scattered by the atmosphere through Rayleigh scattering, thus most space-based satellites including Sentinel-2 do not consider wavebands within the violet section of the EMS (Campbell and Wynne, 2007).

Similarly, the spectral region between 413–420 nm region in the vegetation spectra is influenced by chlorophyll 'a' absorption, whereas the band 600 nm is influenced by chlorophyll 'b' absorption peak (Kumar et al., 2001; Thenkabail et al., 2013). There was a very sharp decline in the reflectance of green vegetation in these regions compared to the Striga classes. This decline is attributed to the absorption of chlorophyll by chlorophyll active plants. On the other hand, the region also demonstrates the low chlorophyll presence and the high influence of the pigments in the flowers (Litchenthaler and Buschmann, 2001).

The spectral differences observed in this study between flowers and green materials were anticipated in our hypothesis to positively produce reliable spectral differences between Striga and co-occurring plants during the Striga flowering period (Ge et al., 2006). From other studies, we know that the waveband 650 nm is the EMS region of vegetation pre-maxima spectral absorption, and 670 nm is sensitive to biomass and leaf canopy (Kumar et al., 2001; Sims and Gamon, 2002; Thenkabail et al., 2013). This was also observed in this study; however, for Striga detection, the significance of spectral absorption at these wavebands was not substantial enough to be useful. Although there was multiple scattering within the NIR caused by water sensitive wavebands, the enormous variation in the magnitude of the reflectance is key in separation of the classes used in this study. GV had the highest reflectance compared to all the other classes in the NIR spectral range.

Apart from chemical composition, vegetation structure can also affect spectral features and influence spectral sensitivity and reflectance (He et al., 2011; Huang et al., 2015). The ability to separate among our studied classes, could also be attributed to flower structure that influences the spectral features associated with the angle and arrangement of the petals. The flower structure coupled with lower water content could have similarly aided in the differences revealed in the prediction of the infestation classes as compared to the pre-flowering Striga which exhibited no difference with GV plots (Ge et al., 2006). Therefore, the best period to predict Striga presence within crop fields is during the peak Striga flowering period. These results concur with Best et al. (1981), who concluded that the best period to discriminate among eight plant species that they were studying was during the flowering and early seed development stages. It is critical to note that using spectral signatures for detecting weeds should be used with caution, since different phenological stages of plants show significant variations in their spectral reflectance depending on the flowering stage of the species (Schmidt and Skidmore, 2003). Carvalho et al. (2013) suggested that further studies might be necessary to analyse what could cause such leaf and flower predictive spectral differences. Additionally, although our results are valid at plot scale, future studies using Sentinel-2 data, airborne or unmanned aerial vehicles (UAV) could be used for seamless wall-to-wall Striga mapping and upscaling from plot scale to field and landscape scales during the peak Striga flowering season.

4.2. Most relevant wavebands and indices using the GRRF approach

Our study employed the robust GRRF for the resampled Sentinel-2 multispectral wavebands, hyperspectral wavebands and their derived VIs variable selection for a multiclass classification. Basically, the traditional RF positively provides the variable importance parameters to lead the GRRF variable selection procedure (Deng and Runger, 2013). Because of the expected very high autocorrelation among the quasi-contiguous hyperspectral wavebands (1-nm interval), the variable importance values were also very similar among the different wavebands. Hence, some of our selected narrow-wavebands for detecting Striga were still autocorrelated. However, the GRRF algorithm decreased the multidimensionality of narrow-wavebands (9 out of 750 were selected) and their derived VIs (6 out of 15 were selected) as well as for the resampled Sentinel-2 multispectral wavebands (6 out of 10) without compromising key information relevant to our Striga and co-occurring vegetation classes. This was in accordance with the findings of Adam et al. (2017); Deng (2013) and Mureriwa et al. (2016) who reported a considerably reduced hyperspectral narrow-waveband dimensionality as a result of using GRRF algorithm. Specifically, our selected narrow-wavebands and VIs concurred and vindicated the importance of the already identified important EMS regions for separating the seven Striga infestation classes. Previously used variable selection methods like "varSelRF" and "Boruta" in R (R Core Team, 2018) are computationally expensive and may yield inexplicit variable importance outputs compared to GRRF (Mureriwa et al., 2016). In other words, the variable importance by-product of such RF-based variable importance procedures could remain dimensionally huge and redundant without identifying a few non-correlated and the most relevant variables.

Furthermore, several plant characteristics such as biochemistry, canopy structure and soil parameters are combined within the canopy spectrum (Große-Stoltenberg et al., 2016). Thus, using canopy-level spectrum rather than leaf-level data is key for further contrast with airborne or satellite remote sensing data such as Sentinel-2 used in the present study. Therefore, in this study we targeted to use VIs that incorporate these characteristics and could easily be upscaled to space-based satellite data. Pigment indices cater for the problem of overlapping absorption characteristics of the different pigments rendering

them more informative than the raw wavebands alone. The fact, our study showed that the most relevant VIs were related to pigments (chlorophylls, carotenoids and anthocyanins) and water content, is attributable to our sampling season which was during the peak Striga flowering and maize crop vegetative growth stage. In addition, the higher chlorophyll content detecting VIs (DDN and Datt4) that were selected among our relevant VIs for Striga detection, are known features that correlated to plant health, leaf area index (LAI) and light use efficiency. The differences in chlorophylls, carotenoids and anthocyanins among our seven Striga classes are vital to the physiological responses and resilience of plants to natural episodic events or seasonal fluctuations (Blackburn, 2007). These dynamics are captured very well in the indices calculated at canopy level. Several studies offer credit to the two-band VIs, which correspond to the flanks of the main chlorophyll absorption feature in the red (530–630 nm) and the waveband located at 700 nm (which resembles waveband 5 of the Sentinel-2 sensor) to be the most sensitive to pigment concentration over the normal range (Kumar et al., 2001).

4.3. Performance of the machine learning classifiers for discriminating among the Striga infestation classes

In general, determination of the best classifier for a particular application case when remotely sensed data are utilized, depends on the accuracy measure selected and the intended objective of the analysis (Maxwell et al., 2018). In our case, we identified the RF and GB as the best classifiers for Striga detection according to their overall accuracies. The resampled Sentinel-2 multispectral wavebands showed good potential to detect and map Striga at acceptable accuracies using the RF and GB algorithm with overall accuracies of 85% and 81%, respectively. However, the RF and GB classifiers experienced challenges by having some false positive cases when attempting to separate the Striga severity classes. These false positive instances are important as detecting a damaging parasitic weed at early stages of invasion is also fundamental for a real time intervention (Große-Stoltenberg et al., 2016). In both cases (i.e. using hyperspectral data or resampled Sentinel-2 multispectral wavebands), we expected the SVM to have performed better than its obtained results in this study, since other studies have shown a high performance of SVM for detecting weed infestation using remotely sensed data (Brereton and Lloyd, 2010; Große-Stoltenberg et al., 2016; Pal and Mather, 2005). However, a relatively small sample size in some of our classes could have hindered the performance of SVM algorithm as it is quite sensitive to imbalanced and small sample sizes (Maxwell et al., 2018). One other reason for a relatively low SVM performance, could be due to the use of a default linear hyperplane and SVM parameters; *viz* gamma (γ) and sigma (C). Studies have shown that optimization of these two SVM parameters would counter for the expected nonlinear relationship among the classes, hence enhancing the performance of the classifier (Abdel-Rahman et al., 2014; Maxwell et al., 2018).

The performance of RF, GB and SVM in this study is in agreement with some recent studies that have utilized leaf-level or canopy-level hyperspectral data and one of these classifiers to detect a plant trait (Große-Stoltenberg et al., 2016; Litchenthaler and Buschmann, 2001; Thenkabail et al., 2013). Specifically, the two non-linear classification algorithms (i.e. GB and RF) attained the best performance results when using the GRRF selected variables for both selected hyperspectral wavebands or when we used the selected resampled Sentinel-2 multispectral wavebands. These results concur with Mureriwa et al. (2016) who used GRRF and RF to detect *Prosopis* using field spectral measurements data and found that reducing the number of redundant spectral variables increased the accuracy of the detection. In all cases, the RF classifier proved to be a very robust and reliable model for predicting subtle differences between classes and non-linear effects from spectral scattering between plant components. This is because RF is robust, yet can still accomplish high prediction accuracies even when the observation data are low or when the variables are highly correlated (James et al., 2017; Thamaga and Dube, 2018). However, GRRF outperforms standard RF in relevant variable selection. It can, therefore, be concluded from our results that, if accurate detection of Striga infestation is to be conducted, combining RF model and GRRF would provide the best model of choice regardless of the dimensions offered by the prediction variables and observations or the mapping scale.

5. Conclusions

In this study, the possibilities of using canopy-level *in-situ* hyperspectral data for predicting the presence and level of Striga infestation using their flowering characteristics are demonstrated. Prediction of seven classes of Striga infestation is possible with satisfactory overall accuracies (up to 94% overall accuracy), specifically during the peak flowering period including at Sentinel-2 spatial and spectral scales. However, due to the reliance on the flowering, it remains a challenge to pre-detect Striga before the damage is done. For more precise results on a global scale, remote sensing could therefore, be used to detect and model the condition of the infested maize rather than targeting Striga itself. The GRRF model provided an easy and accurate variable selection platform that selects a fewer and uncorrelated hyperspectral features relevant to the features of interest like Striga infestation. Specifically, this study shows that the selected narrow-waveband VIs; WBI, ARI, ARI2, Datt4, DDN and PRI.CI2, narrow-wavebands (415 nm, 548 nm, 551 nm, 556 nm, 568 nm, 578 nm, 657 nm, 677 nm and 1060 nm) and Sentinel-2 multispectral wavebands (band 3, band 4, band 9, band 1, band 5 and band 2 in order of importance), are very relevant for Striga infestation prediction in maize fields in semi-arid agro-ecosystem. The machine learning RF classification algorithm emerged as a very robust and reliable model for predicting differences among Striga occurring and other weeds and crops classes. However, there is a need to investigate the temporal and spatial variability of the flowering signal of Striga during the peak flowering season to explore upscaling options for the monitoring of the floral cycle using high spatial resolution multispectral data. Sentinel-2 data coupled with Multiple Endmember Spectral Mixture Analysis (MESMA) which separates spectra within image pixels by identifying the percentage contribution of each class with more than one endmember could also be explored to detect Striga infestation when large scale image data are utilized. This would bring an immense benefit to landscape assessment of the floral cycle and infestation. Findings from this study will be of utmost importance in understanding Striga infestation in heterogenous crop fields in Sub-Saharan Africa. Although the use of the field hyperspectral data in vegetation studies is no longer new, our results indicate the capabilities and application of such remotely sensed data, as a tool for excellent detection of Striga infestation and other vegetation classes. These results provide opportunities to researchers, to apply a similar approach in precision agriculture using airborne or UAV data and platforms to detect the hotspots of Striga infestation at localized scales.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgments

We gratefully acknowledge the financial support for this research by the following organizations and agencies: Biovision Foundation for Ecological Development (Switzerland) and grant number (BV DPP-010/2019); UK's Department for International Development (DFID); Swedish International Development Cooperation Agency (Sida); the Swiss Agency for Development and Cooperation (SDC); Federal Democratic Republic of Ethiopia; and the Kenyan Government. "BTM" was supported by a German Academic Exchange Service (DAAD) In-Region Postgraduate Scholarship. 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.

References

Abdel-Rahman, E.M., Mutanga, O., Adam, E., Ismail, R., 2014. Detecting Sirex noctilio grey-attacked and lightning-struck pine trees using airborne hyperspectral data, random forest and support vector machines classifiers. ISPRS J. Photogramm. Remote Sens. 88, 48–59. https://doi.org/10.1016/j.isprsjprs.2013.11.013.

Abdel-Rahman, E.M., Way, M., Ahmed, F., Ismail, R., Adam, E., 2013. Estimation of thrips (Fulmekiola serrata Kobus) density in sugarcane using leaf-level hyperspectral data. S. Afr. 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.

ASD, 2010. FieldSpec® HandHeld 2 User Manual. pp. 1–140. http://www.geo-informatie.nl/courses/grs60312/material2017/manuals/600860-dHH2Manual.pdf.

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

Best, R.G., Wehde, M.E., Linder, R., 1981. Spectral reflectance of hydrophytes. Remote Sens. Environ. 11, 27–35. https://doi.org/10.1016/0034-4257(81)90004-3.

Blackburn, G.A., 2007. Hyperspectral remote sensing of plant pigments. J. Exp. Bot. 58, 855–867. https://doi.org/10.1093/jxb/erl123.

Breiman, L., 2002. Manual on Setting up, Using, and Understanding Random Forests v3.1. Tech. Report. Stat. Dep. Univ. Calif. Berkeley, pp. 29 http//oz.berkeley.edu/users/breiman https://doi.org/10.2776/85168.

Breiman, L., 2001. Randomforest2001. pp. 1–33. https://link.springer.com/content/pdf/10.1023%2FA%3A1010933404324.pdf.

Brereton, R.G., Lloyd, G.R., 2010. Support vector machines for classification and regression. Analyst 135, 230–267. https://doi.org/10.1039/B918972F.

Campbell, J., Wynne, R., 2007. Introduction to Remote Sensing, 5th ed. Guiford Press, New York Journal of experimental psychology: General.

Carvalho, S., Schlerf, M., van der Puttena, W.H., Skidmore, A.K., 2013. Hyperspectral reflectance of leaves and flowers of an outbreak species discriminates season and successional stage of vegetation. Int. J. Appl. Earth Obs. Geoinf. 24, 32–41. https://doi.org/10.1016/j.jag.2013.01.005.

Chemura, A., Mutanga, O., Dube, T., 2017. Separability of coffee leaf rust infection levels with machine learning methods at Sentinel-2 MSI spectral resolutions. Precis. Agric 18, 859–881. https://link.springer.com/article/10.1007/s11119-016-9495-0.

Chen, J., Shen, M., Zhu, X., Tang, Y., 2009. Indicator of flower status derived from in situ hyperspectral measurement in an alpine meadow on the Tibetan Plateau. Ecol. Indic. 9, 818–823. https://doi.org/10.1016/j.ecolind.2008.09.009.

de Castro, A.I., Torres-Sánchez, J., Peña, J.M., Jiménez-Brenes, F.M., Csillik, O., López-Granados, F., 2018. An automatic random forest-OBIA algorithm for early weed mapping between and within crop rows using UAV imagery. Remote Sens. 10, 1–21. https://doi.org/10.3390/rs10020285.

Deng, H., 2013. Guided Random Forest in the RRF Package. pp. 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.

Dhau, I., Adam, E., Mutanga, O., Ayisi, K., Abdel-Rahman, E.M., Odindi, J., Masocha, M., 2018. Testing the capability of spectral resolution of the new multispectral sensors on detecting the severity of grey leaf spot disease in maize crop. Geocarto Int. 33, 1223–1236. https://doi.org/10.1080/10106049.2017.1343391.

Dube, T., Mutanga, O., 2015. Evaluating the utility of the medium-spatial resolution Landsat 8 multispectral sensor in quantifying aboveground biomass in uMgeni catchment, South Africa. ISPRS J. Photogramm. Remote Sens. 101, 36–46. https://doi.org/10.1016/j.isprsjprs.2014.11.001.

Ejeta, G., Gressel, J., 2007. Intergrating New Technologies for Striga control: Towards Ending the Witch-hunt. World Scientific, Singapore.

Ekeleme, F., Jibrin, J.M., Kamara, A.Y., Oluoch, M., Samndi, A.M., Fagge, A.A., 2014. Assessment of the relationship between soil properties, Striga hermonthica infestation and the on-farm yields of maize in the dry Savannas of Nigeria. Crop Prot. 66, 90–97. https://doi.org/10.1016/j.cropro.2014.09.001.

FAO, I.F.A.D., UNICEF, Wfp, W., 2018. The State of Food Security and Nutrition in the World 2018. Building Climate Resilience for Food Security and Nutrition. FAO, Rome. https://doi.org/10.1093/cjres/rst006. Licence: CC BY-NC-SA 3.0 IGO, Building climate resilience for food security and nutrition.

FieldSpec, 2017. FieldSpec HandHeld2: Light Theory and Measurements Using the Fieldspec HandHeld 2 Portable Spectroradiometer. https://www.malvernpanalytical.com/en/learn/knowledge-center/user-manuals/Copy_of_fieldspec-handheld-2-light-theory-and-measurement.html.

Fisher, R., 1936. The use of multiple measurements in taxonomic problems. Ann. Eugen. 7, 179–188. https://doi.org/10.1111/j.1469-1809.1936.tb02137.x.

Friedman, J., 1999. Stochastic Gradient Boosting. CSIRO C, pp. 1–10 https//statweb.stanford.edu/~jhf/ftp/stobst.pdf.

Ge, S., Everitt, J., Carruthers, R., Gong, P., Anderson, G., 2006. Hyperspectral characteristics of canopy components and structure for phenological assessment of an invasive weed. Environ. Monit. Assess. 120, 109–126. https://doi.org/10.1007/s10661-005-9052-1.

Greenwell, B., Boehmke, B., Cunningham, J., Developers, G., 2019. Gbm: Generalized Boosted Regression Models. R Package Version 2.1.5. CRAN Repos. https//cran.r-project.org/.. .

Große-Stoltenberg, A., Hellmann, C., Werner, C., Oldeland, J., Thiele, J., 2016. Evaluation of continuous VNIR-SWIR spectra versus narrowband hyperspectral indices to discriminate the invasive Acacia longifolia within a mediterranean dune ecosystem. Remote Sens. 8. https://doi.org/10.3390/rs8040334.

He, K.S., Rocchini, D., Neteler, M., Nagendra, H., 2011. Benefits of hyperspectral remote sensing for tracking plant invasions. Divers. Distrib. 17, 381–392. https://doi.org/10.1111/j.1472-4642.2011.00761.x.

Ho, P.-G.P., 2009. Geoscience and Remote Sensing. [WWW Document]. URL (Accessed 2.20.19). https://www.harrisgeospatial.com/Support/Maintenance-Detail/ArtMID/13350/ArticleID/16162/Vegetation-Analysis-Using-Vegetation-Indices-in-ENVI.

Holloway, J., Mengersen, K., 2018. Statistical machine learning methods and remote sensing for sustainable development goals: a review. Remote Sens. 10, 1365. https://doi.org/10.3390/rs10091365.

Huang, J., Wei, Chen, Zhang, Y., Blackburn, G.A., Wang, X., Wei, Chuanwen, Wang, J., 2015. Meta-analysis of the detection of plant pigment concentrations using hyperspectral remotely sensed data. PLoS One 10, 1–26. https://doi.org/10.1371/journal.pone.0137029.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2017. An Introduction to Statistical Learning With Application in R, 8th ed. Springer. Springer, New York. https://doi.org/10.1007/978-1-4614-7138-7.

Jia, K., Wu, B., Tian, Y., Li, Q., Du, X., 2011. Spectral discrimination of opium poppy using field spectrometry. IEEE Trans. Geosci. Remote Sens. 49, 3414–3422. https://doi.org/10.1109/TGRS.2011.2126582.

Karatzoglou, A., Smola, A., Hornik, K., Zeileis, A., 2004. Kenlab—an S4 package for kernel methods in R. J. Stat. Softw. 11, 1–20.

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.

Khan, Z.R., Midega, C.A.O., Hassanali, A., Pickett, J.A., Wadhams, L.J., 2007. Assessment of different legumes for the control of Striga hermonthica in maize and sorghum. Crop Sci. 47, 730–736. https://doi.org/10.2135/cropsci2006.07.0487.

Khan, Z.R., Midega, C.A.O., Pittchar, J.O., Murage, A.W., Birkett, M.A., Bruce, T.J.A., Pickett, J.A., 2014. Achieving food security for one million sub-Saharan African poor through push-pull innovation by 2020. Philos. Trans. R. Soc. B Biol. Sci. 369https://doi.org/10.1098/rstb.2012.0284. 20120284–20120284.

Kuhn, M., Wing, J., Weston, S., Williams, A., Chris, A., Engelhardt Tony, C., Mayer, Z., Kenke, B., The R CoreTeam, I., Michael, B., Reynald, L., Andrew, Z., Luca, S., Yuan, T., 2018. Caret: Classification and Regression Training. R Package Version 6.0-81.

Kumar, A., Manjunath, K.R., Meenakshi, Bala, R., Suda, R.K., Singh, R.D., Panigrahy, S., 2013. Field hyperspectral data analysis for discriminating spectral behavior of tea plantations under various management practices. Int. J. Appl. Earth Obs. Geoinf. 23, 352–359. https://doi.org/10.1016/j.jag.2012.10.006.

Kumar, L., Schmidt, K.S., Dury, S., Skidmore, A.K., 2001. Imaging spectroscopy and vegetation science. In: In: van der Meer, F.D., de Jong, S.M. (Eds.), Imaging Spectrometry: Basic Principles and Prospective Applications, Remote Sensing and Digital Image Processing Vol. 4. Kluwer Academic, Dordrecht, The Netherlands, pp. 111–155.

Landmann, T., Piiroinen, R., Makori, D.M., Abdel-Rahman, E.M., Makau, S., Pellikka, P., Raina, S.K., 2015. Application of hyperspectral remote sensing for flower mapping in African savannas. Remote Sens. Environ. 166, 50–60. https://doi.org/10.1016/j.rse.2015.06.006.

Liaw, A., Wiener, M., Weiner, M., 2002. Classification and regression by random forest. R news 2, 18–22. https://doi.org/10.1177/154405910408300516.

Litchenthaler, H., Buschmann, C., 2001. Chlorophylls and Carotenoids Measurement and UV-vis Characterization. pp. 1–8. https://doi.org/10.1002/0471709085. ch21.

Lukas, W., Lehnert, Hanna, M., Joerg, B., 2018. Hsdar: Manage, Analyse and Simulate Hyperspectral Data in R. R Package Version 0.7.2.

Maxwell, A.E., Warner, T.A., Fang, F., 2018. Implementation of machine-learning classification in remote sensing: an applied review. Int. J. Remote Sens. 39, 2784–2817. https://doi.org/10.1080/01431161.2018.1433343.

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.

Mirik, M., Ansley, R.J., Steddom, K., Jones, D.C., Rush, C.M., Michels, G.J., Elliott, N.C., 2013. Remote distinction of a noxious weed (Musk Thistle: Carduus Nutans) using airborne hyperspectral imagery and the support vector machine classifier. Remote Sens. 5, 612–630. https://doi.org/10.3390/rs5020612.

Mudereri, B.T., Dube, T., Adel-Rahman, E.M., Niassy, S., Kimathi, E., Khan, Z., Landmann, T., 2019. A comparative analysis of PlanetScope and Sentinel-2 space-borne sensors in mapping Striga weed using Guided Regularised Random Forest classification ensemble. ISPRS - Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci XLII-2/W13, 701–708. https://doi.org/10.5194/isprs-archives-XLII-2-W13-701-2019.

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.

Ochungo, P., Veldtman, R., Abdel-Rahman, E.M., Raina, S., Muli, E., Landmann, T., 2019. Multi-sensor mapping of honey bee habitats and fragmentation in agro-ecological

landscapes in Eastern Kenya. Geocarto Int. 0, 1–22. https://doi.org/10.1080/10106049.2019.1629645.
Oswald, A., 2005. Striga control - technologies and their dissemination. Crop Prot. 24, 333–342. https://doi.org/10.1016/j.cropro.2004.09.003.
Pal, M., Mather, P.M., 2005. Support vector machines for classification in remote sensing. Int. J. Remote Sens. 26, 1007–1011. https://doi.org/10.1080/01431160512331314083.
Peña, J.M., Torres-Sánchez, J., de Castro, A.I., Kelly, M., López-Granados, F., 2013. Weed mapping in early-season maize fields using object-based analysis of unmanned aerial vehicle (UAV) images. PLoS One 8, 1–11. https://doi.org/10.1371/journal.pone.0077151.
Qiao, C., Daneshfar, B., Davidson, A.M., 2017. The application of discriminant analysis for mapping cereals and pasture using object-based features. Int. J. Remote Sens. 38, 5546–5568. https://doi.org/10.1080/01431161.2017.1325530.
R Core Team, 2018. R: A Language and Environment for Statistical Computing. URL. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
Rakotoarisoa, M.A., Iafrate, M., Paschali, M., 2012. Why Has Africa Become a Net Food Importer? Explaining Africa Agricultural and Food Trade Deficits, Trade and Market Division. Food And Agriculture Organisation of the United Nations, Rome, FAO.
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.
Rispail, N., Dita, M.A., González-Verdejo, C., Pérez-De-Luque, A., Castillejo, M.A., Prats, E., Román, B., Jorrín, J., Rubiales, D., 2007. Plant resistance to parasitic plants: molecular approaches to an old foe: research review. New Phytol. 173, 703–712. https://doi.org/10.1111/j.1469-8137.2007.01980.x.
Samejima, H., Babiker, A.G., Takikawa, H., Sasaki, M., Sugimoto, Y., 2016. Practicality of the suicidal germination approach for controlling Striga hermonthica. Pest Manage. Sci. 72, 2035–2042. https://doi.org/10.1002/ps.4215.
Sasson, A., 2012. Food security for Africa: an urgent global challenge. Agric. Food Secur. https://doi.org/10.1186/2048-7010-1-2.
Schmidt, K.S., Skidmore, A.K., 2003. Spectral discrimination of vegetation types in a coastal wetland. Remote Sens. Environ. 85, 92–108. https://doi.org/10.1016/S0034-4257(02)00196-7.
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.
Sibanda, M., Mutanga, O., Rouget, M., Odindi, J., 2015. Exploring the potential of *in situ* hyperspectral data and multivariate techniques in discriminating different fertilizer treatments in grasslands. J. Appl. Remote Sens. 9, 096033. https://doi.org/10.1117/1.JRS.9.096033.
Sims, D.A., Gamon, J.A., 2002. Relationships between leaf pigment content and spectral reflectance across a wide range of species. Remote Sens. Environ. 81, 337–354. https://doi.org/10.1016/S0034-4257(02)00010-X.
Smith, A.M., Blackshaw, R.E., 2003. Weed: crop discrimination using remote sensing: a detached leaf experiment. Weed Technol. 17, 811–820. https://doi.org/10.2307/3989767.
Spallek, T., Mutuku, M., Shirasu, K., 2013. The genus Striga: a witch profile. Mol. Plant Pathol. 14, 861–869. https://doi.org/10.1111/mpp.12058.
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. https://doi.org/10.1016/j.rsase.2018.02.005.
Thenkabail, P., Mariotto, I., Gumma, M., Middleton, E.M., Landis, D.R., Huemmrich, F.K., 2013. Selection of hyperspectral narrowbands (HNBs) and composition of hyperspectral twoband vegetation indices (HVIs) for biophysical characterization and. Sel. Top. Appl. EARTH Obs. Remote Sens. 6, 427–439. https://doi.org/10.1109/JSTARS.2013.2252601.
Unachukwu, N.N., Menkir, A., Rabbi, I.Y., Oluoch, M., Muranaka, S., Elzein, A., Odhiambo, G., Farombi, E.O., Gedil, M., 2017. Genetic diversity and population structure of Striga hermonthica populations from Kenya and Nigeria. Weed Res. 57, 293–302. https://doi.org/10.1111/wre.12260.
Vapnik, V., 1979. Estimation of Dependences Based on Empirical Data. Nauk. Moscow, Transl, vol. 27. Springer Verlag, New York, pp. 5165–5184 1982.
Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics With S, fourth edition. Springer, New York ISBN 0-387-95457-0.