

A two-step approach for detecting Striga in a complex agroecological system using Sentinel-2 data

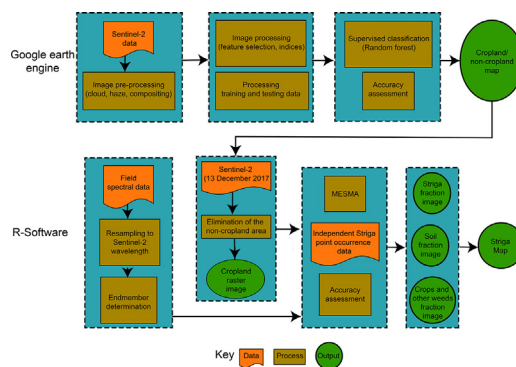
Bester Tawona Mudereri ^{a,b,*}, Elfatih Mohamed Abdel-Rahman ^{a,c}, Timothy Dube ^b, Saliou Niassy ^a, Zeyaur Khan ^a, Henri E.Z. Tonnang ^a, Tobias Landmann ^{a,d}

^a International Centre of Insect Physiology and Ecology (icipe), P. O. Box 30772, 00100, Nairobi, Kenya
^b Department of Earth Sciences, University of the 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

HIGHLIGHTS

- African cropping systems hinder the spaceborne detection of Striga weeds
- Remotely sensed spectral data often shadows analytical spectral characteristics
- A cropland mask was used as fraction constraint in MESMA to enhance Striga detection
- This integration successfully reduced spectral confusion and computation time
- The method is plausible for Striga detection in complex agroecological systems

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
 Received 15 May 2020
 Received in revised form 1 July 2020
 Accepted 5 October 2020
 Available online 21 October 2020

Editor: Manuel Esteban Lucas-Borja

Keywords:
 Africa, croplands
 Endmember selection
 Google Earth Engine
 Invasive weeds
 Spectral mixture modeling

ABSTRACT

Information on weed occurrence within croplands is vital but is often unavailable to support weeding practices and improve cropland productivity assessments. To date, few studies have been conducted to estimate and map weed abundances within agroecological systems from spaceborne images over wide-area landscapes, particularly for the genus *Striga*. Therefore, this study attempts to increase the detection capacity of *Striga* at subpixel size using spaceborne high-resolution imagery. In this study, a two-step classification approach was used to detect *Striga* (*Striga hermonthica*) weed occurrence within croplands in Rongo, Kenya. Firstly, multitemporal and multi-year Sentinel-2 (S2) data (2017 to 2018) were utilized to map cropland and non-cropland areas using the random forest algorithm within the Google Earth Engine. The non-cropland class was thereafter masked out from a single date S2 image of the 13th of December 2017. The remaining cropland area was then used in a subpixel multiple endmember spectral mixture analysis (MESMA) to detect *Striga* occurrence and infestation using endmembers (EMs) obtained from the in-situ hyperspectral data. The gathered in-situ hyperspectral data were resampled to the spectral waveband configurations of S2 and three representative EMs were inferred, namely: (1) *Striga*, (2) crop and other weeds, and (3) soil. Overall classification accuracies of 88% and 78% for the pixel-based cropland mapping and subpixel *Striga* detection were achieved, respectively. Furthermore, an F-score (0.84) and a root mean square error (0.0075) showed that the MESMA subpixel algorithm provides plausible results for predicting the relative abundance of *Striga* within each S2 pixel at a landscape scale. The capability

* Corresponding author at: International Centre of Insect Physiology and Ecology (icipe), P. O. Box 30772, 00100, Nairobi, Kenya.
 E-mail address: bmudereri@icipe.org (B.T. Mudereri).

ity of MESMA together with a cropland classification hierarchical approach was thus proven to be suited for Striga detection in a heterogeneous agroecological system. These results can be used to guide in the adaptation, mitigation, and remediation of already infested areas, thereby avoiding further Striga infestation of new croplands.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

There is a great need for precise information on the occurrence of harmful weeds within croplands to enhance scaling out of interventions and mitigation of crop production constraints, that may impede agricultural products' supply chains (Mutanga et al., 2017). Striga is a parasitic weed of global economic importance that penetrates the roots of the cereal crops after germination and withdraws nutrients and water, thereby suffocating and out-competing the host cereal crops (Hassanali et al., 2008). Consequently, the productivity of economically fundamental host crops like maize, sorghum, and rice can deteriorate considerably, causing extensive grain yield reductions and unprecedented socioeconomic impacts (Ejeta and Gressel, 2007). Striga weed detection using remote sensing techniques such as the in-situ hyperspectral methods (Mudereri et al., 2020) or high-to-medium spatial resolution satellite data (i.e. PlanetScope and Sentinel-2 (S2)) are essential tools to help identify weed distributional patterns and priority areas at plot and/or field scales (Mudereri et al., 2019a). Spectral responses in small scale agroecological systems, typical for Africa, exhibit a large intra- and inter-field variability which results in detection errors when discriminating between co-occurring crops and weeds (i.e. maize and Striga) (Mudereri et al., 2019a). Specifically, the heterogeneity, fragmentation, and complex crop cycles (Xiong et al., 2017a) of African cropping systems hinder the detection of hidden spectral properties of the Striga weed, when pixel-based detection approaches are employed.

On the other hand, the complex structure of remotely sensed spectral information often shadows analytical spectral characteristics and obstructs the comprehensive characterization of targeted constraints within the croplands (Somers et al., 2011). However, studies have shown that the mixed pixels (e.g. crop and Striga) can be decomposed using subpixel classification methods i.e. spectral mixture analysis (SMA: Powell et al., 2007; Huovinen et al., 2020). Numerous variants of SMA to model subpixel fractions have been used by other studies such as the simple spectral mixture analysis (sSMA: Somers et al., 2011), Monte Carlo spectral mixture model (AutoMCU: Asner and Lobell, 2000), Bayesian spectral mixture analysis (BSMA: Song, 2005), linear mixture analysis (Heinz and Chang, 2001) and multiple endmember spectral mixture analysis (MESMA: Roberts et al., 1998). These methods are all set and operationalized by: (1) determining how many pure spectra known as endmembers (EMs) are present in an image data, (2) identifying the biophysical nature of each of those EMs within a pixel, and (3) estimating the fractional abundances of each EM in a pixel (Correa Pabón et al., 2019; Somers et al., 2011).

MESMA is, by far, the most extensively used SMA method that yields a relatively low classification error and provides high detection accuracies when compared to the other SMA methods (Degerickx et al., 2019; Wang et al., 2014). The MESMA approach has been used in a wide range of applications including characterization of urban environments (Franke et al., 2009), mapping of fire (Quintano et al., 2013), plant species (Roberts et al., 1998), marshes (Li et al., 2005; Rosso et al., 2005) and agricultural agroecosystems (Njenga, 2016), among others. However, no or a few studies have yet to use MESMA-based fraction images to estimate Striga (*Striga hermonthica*) occurrence or severity. Thus, this study is the first attempt to test the efficacy of MESMA in estimating and mapping Striga occurrence and fractions within croplands in Africa using a test site in Kenya.

To effectively perform a subpixel classification method, Degerickx et al. (2019) suggested using a-priori classification or segmentation

along with SMA approaches to constrain EMs to certain segments of the image to ensure a reduction in the spectral confusion. To detect invasive weeds like Striga that infest croplands using image data and subpixel SMA, one would need to first classify the croplands, before employing any SMA like MESMA. Mapping of croplands using single-date or multi-date multispectral images and machine learning classification algorithms is well documented in the literature (Belgiu and Csillik, 2018; Gumma et al., 2019). Among the machine learning algorithms used in most of these studies, the random forest (RF) performs relatively better than the other methods for delineating croplands (Belgiu and Csillik, 2018; Immitzer et al., 2016).

In this study, the Google Earth Engine (GEE) was used to execute accurate and localized cropland mapping. The strength and versatility of the GEE and the subpixel MESMA to manipulate S2 multitemporal data were exploited to execute a two-step cropland and Striga weed classification approach. This approach was necessitated by the need to constrain the classification of Striga occurrence within the cropland area to considerably reduce the number of EMs. Various studies in agriculture monitoring and management have used the GEE platform to add and curate their data while exploiting Google's cloud resources to undertake all the processing procedures (Kelley et al., 2018; Landmann et al., 2019). The objective of this study was thus to advance the feasibility of Striga weed detection using the subpixel MESMA within croplands derived using the high resolution (10 m) freely available multispectral S2 images and their respectively derived vegetation indices (VIs) in the small-holder farming system.

2. Study area

The study was conducted in the Rongo sub-county (coordinates: 0° 39' 12" S; 34° 35' 40" E and 0° 59' 16" S; 34° 37' 21" E), covering an area of ~213 km². Rongo is in the Migori county of western Kenya, occurring at an altitude of 1470 m (Fig. 1). A tropical and subtropical climate characterizes the study area with a bimodal rainfall distribution. Average annual rainfall of 1600 mm across the two rainy seasons i.e. during 'short rains' season spanning November to January and 'long rains' season occurring between March and June characterizes the study area. The study area experiences relative humidity that ranges between 50% and 70% throughout the year, while the annual average temperature is 20.6 °C.

In Rongo, the savanna grassland biome dominates the natural ecosystems, often flanked with deciduous and exotic forest vegetation. On the other hand, the agricultural cropping system in the study area is mainly mixed small-scale farms, with an average field size of 0.1 ha. Maize, bean, cassava, green gram, groundnut, and fruits such as avocado, banana, pawpaw, mango, and indigenous vegetables as food crops are common in Rongo for predominantly subsistence purposes. Sugarcane grown mainly in the southern region of the study area is the main cash crop, among others. Rainfall variability, insect pests, and the invasive Striga weed are the major crop production constraints in the study area.

3. Methods

3.1. Two-step hierarchical classification approach

The proposed methodology uses a two-step hierarchical approach to detect the occurrence of Striga i.e. (1) S2 time-series composites to detect and distinguish cropland and non-cropland over a series of agricultural growing seasons using RF classifier in a semi-automatic approach

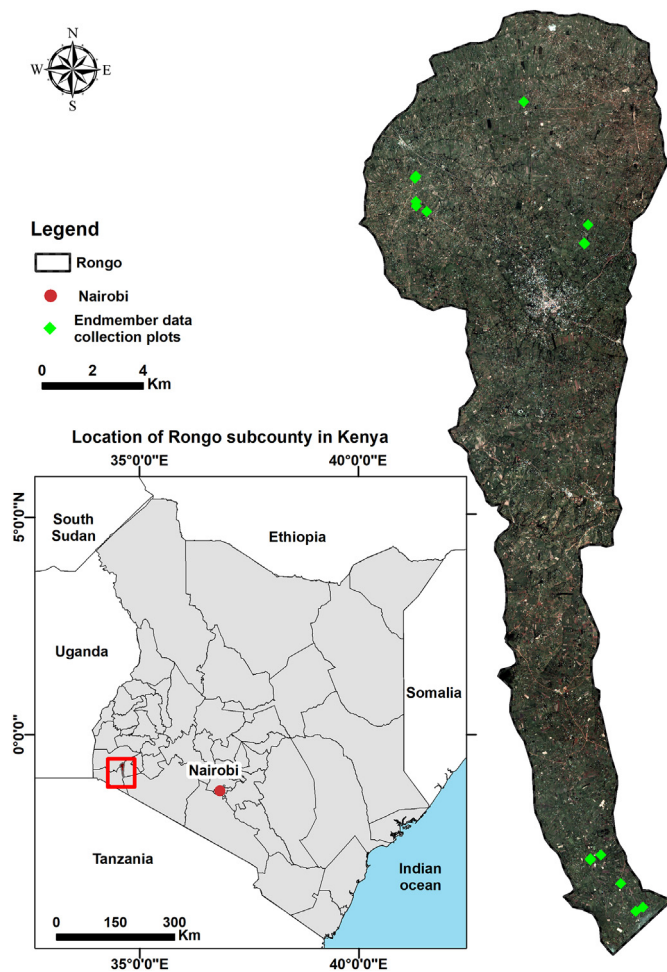


Fig. 1. The location of Rongo in Kenya. The light green diamonds show the location where the spectral endmembers were extracted. The background layer is the Sentinel-2 image of the 13th of December 2017 displayed in the true colour using the red, green, and blue (RGB) band combination of Sentinel-2 bands 4,3, and 2, respectively.

in GEE and (2) spectral unmixing of the derived cropland using a single-date (13th December 2017) S2 image and MESMA in R-software to assess the occurrence of *Striga*. The generalized workflow of the classification approach used in this study is shown in Fig. 2. The details of each step are thereafter described in the subsequent sub-sections.

3.2. *Striga* and land use/ cover (LULC) field data collection

Striga 'presence' ($n = 52$: maize fields with at least one visible stand of *Striga* up to ∞ stands) or 'absence' ($n = 20$: maize fields with no visible stands of *Striga*), together with other cropland data were obtained from field surveys conducted between the 12th and 16th of December 2017, which coincided with the peak *Striga* period and the maximum phenological stage of the maize in the study area. A global positioning system (GPS) instrument (± 3 m accuracy) was used to locate the field *Striga* reference data. Also, reference data on different land use and land cover (LULC) classes i.e. bare land (including fallow and abandoned cropland), built-up area, natural vegetation (forests, shrubland, and grasslands), and water were collected through digitizing regions of interest (ROI) on Google Earth® (Chemura et al., 2017) and are summarized in Table 1. The administrative ward boundaries were used as strata in a stratified random sampling approach, to determine areas of *Striga* 'presence' and 'absence', other LULC classes and to reduce sampling bias. The *Striga* and LULC classes were arranged into two main

classes for classification i.e. cropland and non-cropland to create a crop mask. A total of 260 sample polygons consisting of 5890 pixels for the two classes were derived and divided into 77% for image classification training and 23% for testing the accuracy of the classification (Table 1). To perform a pixel-based classification, the sample points at a pixel scale were organized in a Google fusion table and retrieved in the GEE, then the corresponding input sample pixel values were extracted at the S2 image pixel resolution (i.e. 10 m).

3.3. Cropland mask generation in the GEE

The use of the GEE in this study was employed to map cropland and generate a cropland mask to reduce the propagation of classification error. The freely accessible cropland maps are derived over large areas at coarse spatial resolution hence have many limitations such as inability to account for fragmentation, failure to delineate small farms and to adequately capture the influence of mixed cropping systems, which leads to confusion with other land cover types (Oliphant et al., 2019). As a result, such cropland maps are inadequate and not useful in areas where crop fields are small (< 0.1 ha) and within mixed cropping systems such as those in the small-holder farming sector which characterize the present study area in Kenya.

3.3.1. Sentinel-2 (S2) data compositing

S2 level 1C images, which were provided at the top of atmosphere reflectance, were sourced from the GEE. Image processing and analysis for generating the non-cropland mask was also implemented in the GEE. The procedures executed in GEE included image cloud filtering, atmospheric correction (converted to the bottom of atmosphere reflectance) and normalization for illumination effects (i.e. shade) using median compositing, spatial resampling to the 10 m pixel size (for bands whose pixel size are > 10 m), vegetation indices calculation, machine learning classifier parameterization, creating the final cropland classification map, and assessing the accuracies of the classification product. A relatively cloud-free layer stack image input was generated by stacking the S2 reflectance image collection ($n = 121$) within two consecutive years (i.e. between 1st of January 2017 and 31st of December 2018) and then applying a cloud and shadow mask. Pixel-based image compositing is a common procedure to condense the number of pixels with redundant and invalid data due to the atmospheric interference, shadow, or other noise remaining after pre-processing (Bey et al., 2020). Clouds and shadows cause radiometric distortions that hamper the operational use and product generation by compromising the accurate estimation of biophysical parameters (Wu et al., 2018). Thus, the elimination of contaminated pixels is an important procedure that guarantees a reduction in spectral error.

The median compositing method has been reported to be computationally and technically less demanding and provides the best results when compared to other pixel-based image compositing methods such as (i) maximum ratio value, (ii) annual greenness pixel, (iii) best pixel based on the distance to the nearest cloud, and (iv) seasonal greenest pixel (Bey et al., 2020). Use of the median value eliminated clouds (high pixel values) and shadows (low pixel values) from each band, over time (Bey et al., 2020; Gumma et al., 2019; Xiong et al., 2017b). Using the 'median ()' function in GEE, the median values for all the 121 S2 images were used for each band to capture the variability among seasons and to select the most representative pixel values over the entire period.

Five vegetation indices (VIs) namely normalized difference vegetation index (NDVI: Rouse et al., 1974), modified soil adjusted vegetation index (MSAVI: Qi et al., 1994), enhanced vegetation index (EVI: Huete et al., 2002), and two red-edge (RE) vegetation indices i.e. RE-NDVI (Sibanda et al., 2019) and RE-EVI were derived from the bands of S2 at 10 m pixel size and added to the spectral band composite. The two RE band-indices were computed after resampling the respective bands to the 10 m pixel size. This was done to match the spatial

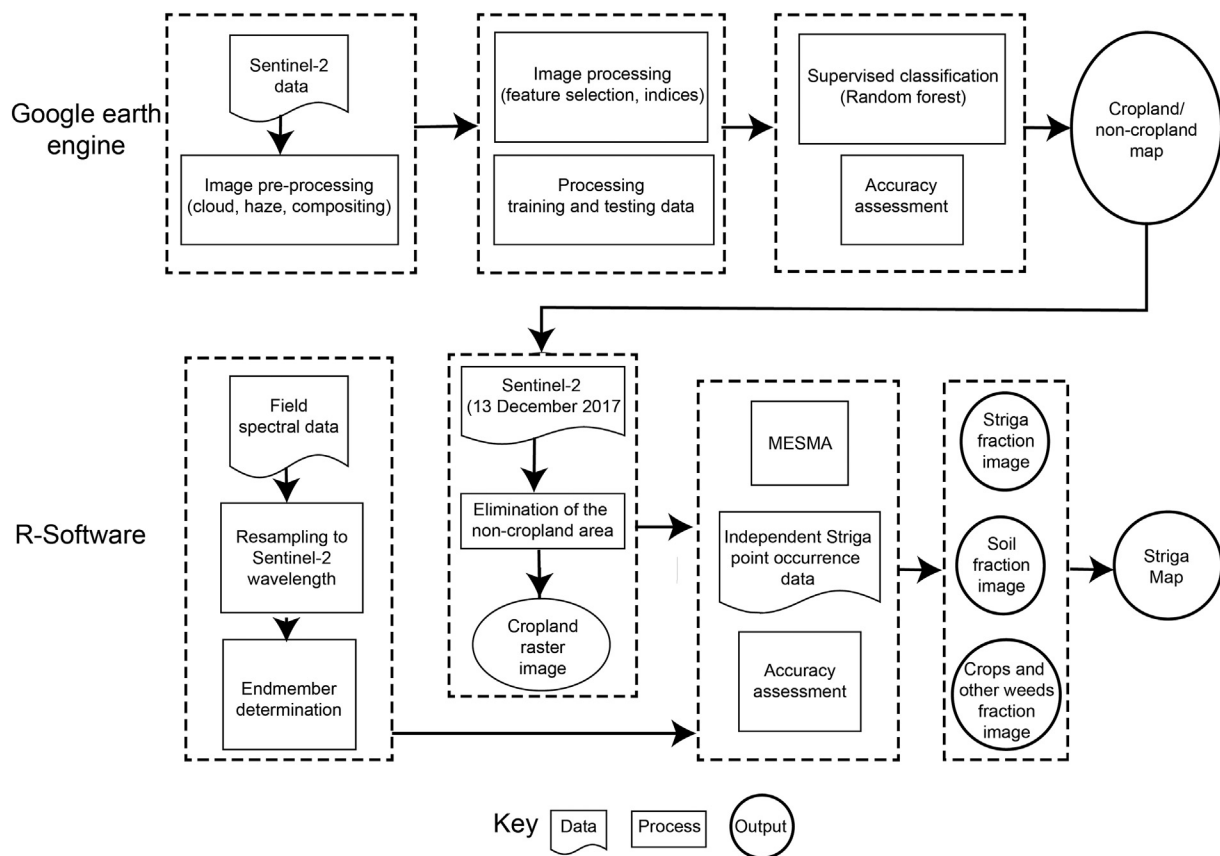


Fig. 2. The general workflow of the two-step hierarchical approach for detecting the Striga weed.

resolution of the visible (blue, green, and red) and NIR bands. These indices were selected as they are designed to capture the sensitivity of vegetation features while minimizing the influence of the confounding factors such as atmospheric effects and soil reflectance (Bannari et al., 1995; Xue and Su, 2017). These indices have also been reported by other studies as the best in capturing vegetation variabilities over time (Chemura et al., 2017; Sibanda et al., 2019). These indices are calculated as shown in Eqs. (1)–(5):

$$NDVI = \frac{\rho_{NIR} - \rho_{Red}}{\rho_{NIR} + \rho_{Red}} \tag{1}$$

$$MSAVI = \frac{2\rho_{NIR} + 1 - \sqrt{(2\rho_{NIR} + 1)^2 - 8(\rho_{NIR} - \rho_{Red})}}{2} \tag{2}$$

$$EVI = 2.5 \times \frac{\rho_{NIR} - \rho_{Red}}{\rho_{NIR} + 6 \times \rho_{Red} - 7.5 \times \rho_{Blue} + 1} \tag{3}$$

where ρ_{NIR} , ρ_{Red} , ρ_{Blue} in Eqs. (1), (2), and (3) represent the near-infrared red and blue reflectance values, respectively for a given pixel.

$$RE-NDVI = \frac{\rho_{NIR} - \rho_{RE2}}{\rho_{NIR} + \rho_{RE2}} \tag{4}$$

$$RE-EVI = 2.5 \times \frac{\rho_{NIR} - \rho_{RE2}}{\rho_{NIR} + 6 \times \rho_{RE2} - 7.5 \times \rho_{Blue} + 1} \tag{5}$$

where ρ_{RE2} in Eqs. (4) and (5) represent S2 band 6 or red-edge 2 (RE2) reflectance values for each given pixel.

Table 1
Striga and Land use/ cover (LULC) classes and samples sizes used in the classification of the cropland and Striga in the Rongo study area, Kenya.

Striga/ LULC class	Description	Cropland class	Sample polygon	Train pixels	Test pixels
Striga infested maize fields	Crop fields with Striga	Cropland	52		
Non-Striga infested maize fields	Crop fields without Striga	Cropland	20		
Other crops	Other non-maize cropland	Cropland	46		
		Total	118	2362	720
Built-up	Human-made constructions	Non-cropland	52		
Bare land	Surfaces without vegetation	Non-cropland	21		
Natural vegetation	Woodland and grassland	Non-cropland	35		
Water	Water bodies	Non-cropland	34		
		Total	142	2158	650
Total			260	4520	1370

3.3.2. Random forest classification

The RF algorithm (Breiman, 2001) was used to classify the cropland and non-cropland area in GEE using the median composite reflectance of the S2 wavebands and the additional five VIs. RF is an ensemble machine learning classifier that has accomplished effective classification and prediction results in many remote sensing studies, including cropland mapping (Belgiu and Csillik, 2018; Oliphant et al., 2019). A detailed explanation of RF and its efficiency in remote sensing is provided in Abdel-Rahman et al. (2014). RF builds compound decision trees (*n*tree) in a machine learning ensemble algorithm approach for classification and regression (Breiman, 2001). For each of the decision trees, a bootstrap sample (2/3 of the original data referred to as “in bag” data) is grown. These randomly sampled subsets are used to split several nodes of these decision trees using random subsets of variables for classification (*m*try) with the default *m*try value calculated as the square root of the total number of variables (Abdel-Rahman et al., 2014). The class with the majority votes from all the generated trees is then provided as the final class prediction (e.g. cropland or non-cropland) (Mudereri et al., 2019b). In this study, the default *n*tree and *m*try settings, viz. 500 trees and 3 *m*try were used to perform the pixel-based RF classification for cropland mapping. On the other hand, the algorithm was trained using the independent 77%-pixel samples (4520) as shown in Table 1. The non-cropland class was masked out from the S2 image of the 13th of December 2017 and the remaining cropland class was then used to perform MESMA as described in Section 3.4.

3.4. Endmember selection and collection

The most vital step to a successful and valid SMA encompasses determining the number, type, and matching spectral signatures of EMs (Song, 2005; Somers et al., 2011). These spectral signatures of the EMs are obtainable either directly from the satellite image data reference spectra or spectral libraries, and portable spectroradiometers (Landmann, 2003; Somers et al., 2011).

In this study, three EM spectra within the cropland class were identified and used in the MESMA namely: (1) flowering Striga (2) crop and other weeds, and (3) soil. These three EMs were collected using canopy-level in-situ hyperspectral data collected with the ASD FieldSpec® Handheld 2™ spectroradiometer (HH2; ASD, 2010). We purposively sampled 14 maize field plots (30 m × 30 m each) during the peak Striga flowering window (12th–16th December 2017). In each plot, five quadrats measuring 1 m × 1 m each were identified 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 the other quadrat was placed at the center of the sampling plot (i.e. 0.25 m from each side of a quadrat). The 1 m × 1 m quadrats were used as a guide to locate the center of the focus area for the in-situ spectral measurements of the EMs (Fig. 3a).

Within each quadrat, the ASD with a ground focal area of about 0.19 m² was used to collect the EM spectral data (Fig. 3b). The ASD instrument used in this study acquires reflected radiation in 325–1075 nm of the electromagnetic spectrum with a built-in 2 nm sampling resolution (ASD, 2010). The collected spectra are automatically resampled to a 1 nm spectral range. A bare optical input at a nadir full conical angle field of view (25°) was employed at 1 m above the maize crop to collect the data. This setting allows the instrument to capture spectra in an area of ~0.5 m in diameter on the target (FieldSpec, 2017; Sibanda et al., 2015). This target area unit was enough for capturing the precise spectral signal of the three EMs i.e. Striga, crops with other weeds, and soil. A full description of the EM data collection and how the guiding quadrats were laid out are provided in Mudereri et al. (2020).

A total of 120 spectra samples were collected and grouped into the three target endmembers i.e. (1) flowering Striga (*n* = 40), (2) crop and other weeds (*n* = 40), and (3) soil (*n* = 40). These samples were filtered using the ‘noiseFiltering’ function and smoothed using the

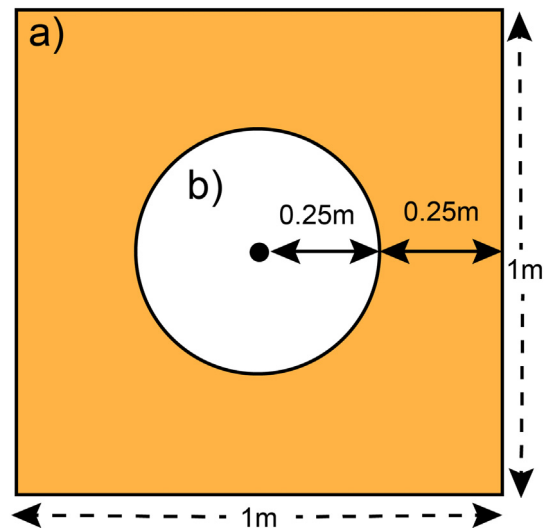


Fig. 3. a) The guiding quadrat used to estimate the Striga infestation level and b) the ground focal area of the analytical spectral device (ASD) measurements.

‘Savitzky–Golay’ filter in R software (R Core Team, 2020) using the ‘hsdar’ package (Lukas et al., 2018). The spectral resampling to the configuration of S2 was subsequently conducted using the spectral response function, i.e. ‘SpectralResampling’ in the ‘hsdar’ package. The average values corresponding to each band were then used as the input to the EM values for running the MESMA in R software (R Core Team, 2020).

3.5. Striga detection using multiple endmember spectral mixture analysis (MESMA)

The cropland area of the S2 reflectance image of the 13th of December 2017 that was used in the MESMA was pre-processed in the GEE as described in Section 3.3.1 of this study. The choice of the image sensing date was influenced by the coincidence with the period of field reference data collection, the highest maize phenology, peak Striga flowering, and the optimal time for Striga separability from the co-occurring crops and other weeds (Mudereri et al., 2020). The MESMA models the spectral data as linear combinations of pure spectra, called EMs while allowing the types and number of EMs to vary on a per-pixel basis (Roberts et al., 1998). Thus, each class is characterized by unique sets of EMs and their fractions (Thenkabail et al., 2019). For each pixel, MESMA iteratively runs several candidate models (Frank et al., 2009) and the best model is assessed based on pixel fractions identified, their residuals and having the smallest root mean square error (RMSE) when compared to the spectral curve of the pixel (Degerickx et al., 2019; Thenkabail et al., 2019). The models are thus applied to the entire image on a per-pixel basis. In this study, the MESMA algorithm within the ‘RStoolbox’ package (Leutner et al., 2019) in R-software (R Core Team, 2020) was used. The ‘RStoolbox’ package for MESMA uses the non-negative least squares (NNLS) regression through a sequential coordinate-wise algorithm. The MESMA was executed using the default settings as follows: method NNLS, 400 iterations, and tolerance of 1×10^{-9} . The outputs from the MESMA algorithm are individual bands representing the estimated pixel fraction and probability of occurrence of each of the tested EMs (0–1) and RMSE per pixel. The base equations of MESMA and RMSE are given by Eqs. (6)–(8) (Thenkabail et al., 2019):

$$\rho^\lambda = \sum_{i=1}^N \hat{f}_i \rho_{i\lambda} + \epsilon_\lambda \quad (6)$$

and

$$\sum_{i=1}^N f_i = 1 \tag{7}$$

$$RMSE = \sqrt{\frac{\sum_{\lambda=1}^M (\epsilon_{\lambda})^2}{M}} \tag{8}$$

where ρ'_{λ} is the reflectance at a given wavelength (λ), $\rho_{i\lambda}$ is the reflectance of EM i at λ , f_i is the cover fraction of EM i , N is the total number of EMs in the model, ϵ_{λ} is the residual error of the model, and M is the total number of bands.

3.6. Accuracy assessment

Accuracy assessment and validation are a key component of any thematic map production, particularly when using remotely sensed data (Dube et al., 2019). The process demands randomly sampled high-quality reference data sets enumerated at suitable spatial and temporal scales (Mupfiga et al., 2016). The classification results of each of the hierarchical steps i.e. cropland and Striga maps were compared to random samples of validation pixels to assess the classification accuracy. Due to the unavailability of percentage coverage reference observations for Striga, soil, and other crops and weeds classes to compare with the predicted class fractional coverage, the accuracy of MESMA results was validated using the standard and traditional classification confusion matrix method. A pixel in the MESMA map was classified as the material (i.e. Striga, soil, or crops with other weeds) with the highest fractional coverage. The soil and crops with other weeds classes were then combined as non-Striga.

Specifically, the performance of the RF classifier used in cropland mapping and MESMA were validated using the accuracy assessment metrics, derived from the respective confusion matrices (i.e. overall accuracy (OA), user's accuracy (UA), producer's accuracy (PA), and the F-score) and the RMSE. The OA was calculated by dividing the total correctly classified instances by the total samples considered in the classification (Quintano et al., 2013). On the other hand, the PA (i.e. error of omission) was derived from dividing the correctly classified instances in each class by the total number of reference samples for that class, while UA (i.e. error of commission) was obtained by dividing the correctly classified instances in each class with the total number of instances that were classified to belong to that class (Kyalo et al., 2017). The F-score evaluates the accuracy of a class using the precision (positively classified values) and recall (the number of relevant instances that were actually classified, also called sensitivity) (Graesser and Ramankutty, 2017; Kyalo et al., 2017). For each of the mapping approaches (i.e. cropland mapping in GEE and Striga occurrence mapping using MESMA), the error matrices that provided all the four metrics i.e. OA, UA, PA, and F-score were established.

4. Results

4.1. Cropland mapping

Approximately 54% (114.3 km²) of the Rongo area was classified as cropland while the remaining area of 98.7 km² (46%) was non-cropland. Remarkably, the forest areas, roads, bare land, built-up areas, and their boundaries were accurately detected. The results indicate that the croplands in Rongo are diverse and irregular in shape, direction, and sizes (Fig. 4). Due to the inter-annual variability, some of

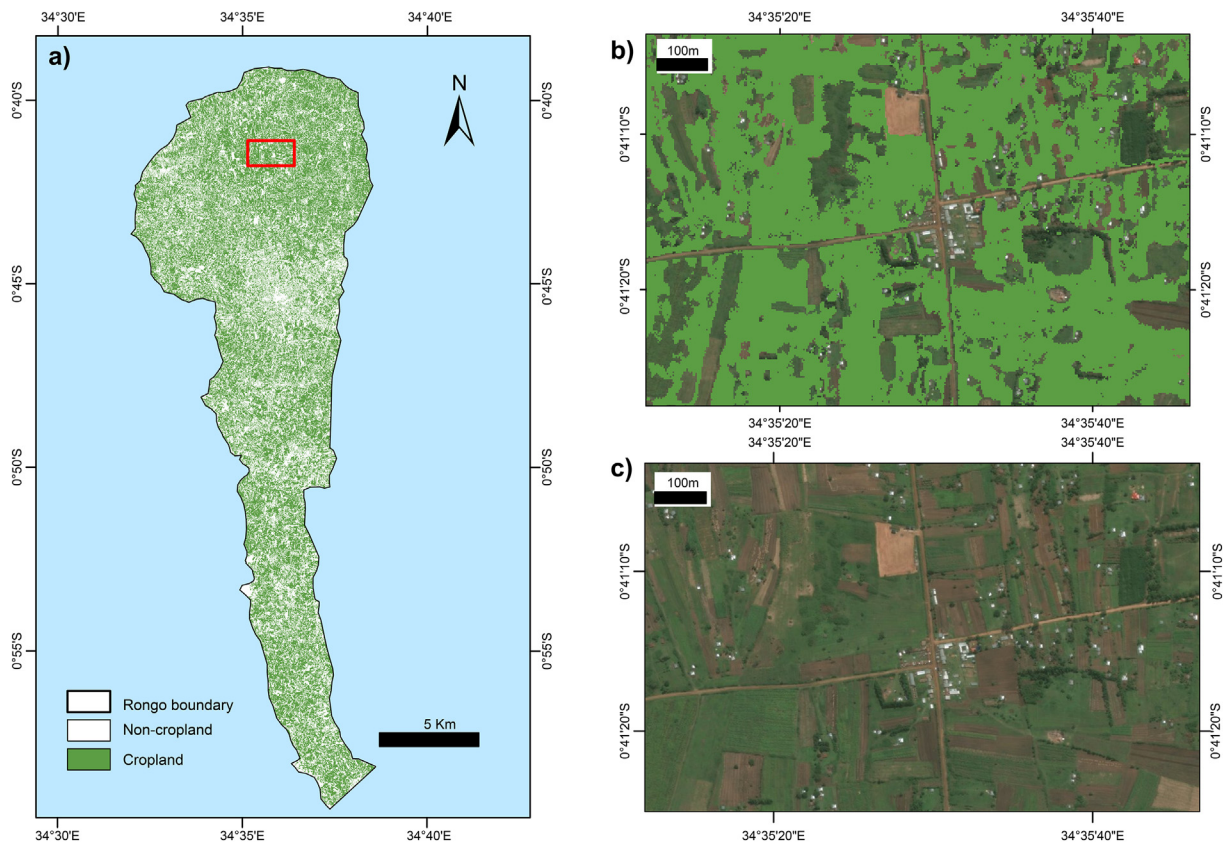


Fig. 4. Visual comparison of the 10 m cropland map extents derived using Sentinel-2 data and random forest classifier in GEE a) Rongo sub-county, b) a subset cropland extent shows crop fields in light green colour overlaid on a Google Earth image, and c) a reference Google Earth image zoomed-out from the red rectangle in a).

Table 2
Classification confusion matrix for the cropland and non-cropland classes mapped in the Rongo sub-county using Sentinel-2 data and random forest classifier in GEE.

Reference data		Cropland	Non-cropland	Total	UA (%)
Map data	Cropland	527	143	670	79
	Non-cropland	24	676	700	97
	Total	551	819	1370	
	PA (%)	96	83		
	F-score	0.87	0.89		
OA = 88%					

the croplands were fallow or probably abandoned as shown from the Google high-resolution imagery. However, the classified image managed to capture the inter-seasonal variability.

Table 2 shows that the overall cropland classification accuracy was comparatively high (88%) with F-scores closer to 1 (0.87 and 0.89 for the cropland and non-cropland classes, respectively). Table 2 also shows that the error of omission for the cropland class was 4% (PA = 96%) and the error of commission was 21% (UA = 79%), while for non-cropland the error of omission was 17% (PA = 83%) and error of commission was 3% (UA = 97%).

4.2. Endmembers used in Striga detection using MESMA

Fig. 5 demonstrates the variation in spectral responses of the three selected EMs within the 8 resampled S2 bands that correspond with the wavelength range of the ASD instrument used. The RE2, RE3, NIR, and NIRn revealed spectrally distinguishable differences in the EM reflectance values i.e. ± 0.3 emanating from the different compositions of the classes. However, there is no substantial difference in the EM values within the same class (e.g. Striga EM) across the four stated bands (i.e. RE2, RE3, NIR, and NIRn). On the other hand, the visible bands (blue, green, and red) and the RE1 did not show substantial differences for both within classes and across the three EMs (Fig. 5).

The Striga EM demonstrated the influence of the Striga flower compaction and colour by having the highest EM value in the red and higher than the crops and other weeds EM in RE1. In contrast, the soil EM shows higher reflectance values in the blue and red spectral bands.

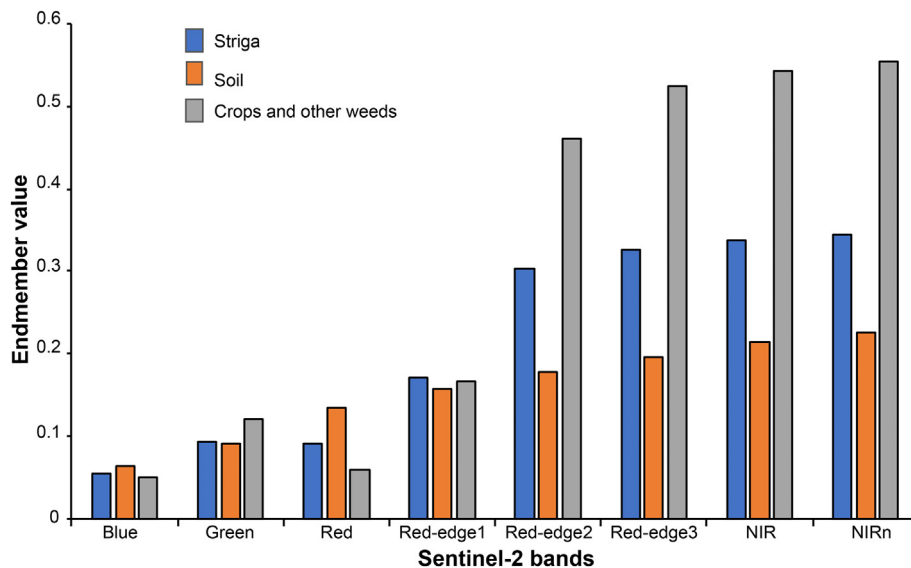


Fig. 5. Three endmembers and their respective values derived from the resampled eight Sentinel-2 spectral bands used in the multiple endmember spectral mixture analysis (MESMA) for Striga detection. The eight bands correspond to the blue, green, red, RE1, RE2, RE3, NIR, and NIRn waveband areas.

The crops and other weeds EM dominated the other five bands (i.e. green, RE2, RE3, NIR, and NIRn).

4.3. Frequency of pixel fractions of the three endmembers

Fig. 6 shows the distribution and frequency of pixel fractions across the three EMs tested in this study. The crops and other weeds EM dominated most of the pixels compared to those of Striga and soil EMs in the study area. Although many of the pixels and the larger pixel fractions were occupied by crops and other weeds, results show fewer pixels that had crop and other weeds fractions that are >0.6. The huge density of Striga pixels was between 0.1 and 0.4 pixel fractions with >250,000 pixels having zero fractions of Striga. On the contrary, there were very few pixels that exhibited close to zero value of the crops and other weeds fractions in the entire study area. The proportion of soil EM, when compared to the other two EMs, was very low as shown by the few pixels representing the soil fraction and very low soil fractions within those few available soil representative pixels. Most of the pixels had a low RMSE (< 0.01) and a maximum RMSE of 0.024 across all the pixels was obtained. Remarkably, this RMSE value represents almost 100% of the whole study area. The average RMSE (0.0075) revealed that there were generally low EM prediction errors using MESMA for the entire study area.

MESMA generated four fraction images for the three EMs and their corresponding RMSE. Fig. 7 shows the classification results from the MESMA of the fraction images of the three EMs. The results showed that the RMSE was generally low across the entire study area, except for the central areas of Rongo. Striga occurrence was high in the north and northwestern sides of Rongo as visually shown by the red sub-pixels in Fig. 7.

4.4. Striga infestation in Rongo sub-country

The false-colour image of the three EMs predicting the pixel fraction using MESMA predicted that high proportions of Striga occur in the northwestern region of Rongo (Fig. 8b), while in the southern region it is generally low (Fig. 8c). Interestingly, Fig. 8 shows that some fields were completely free of Striga infestation while in some other fields the entire field (~ 30 m × 30 m) was completely occupied by Striga.

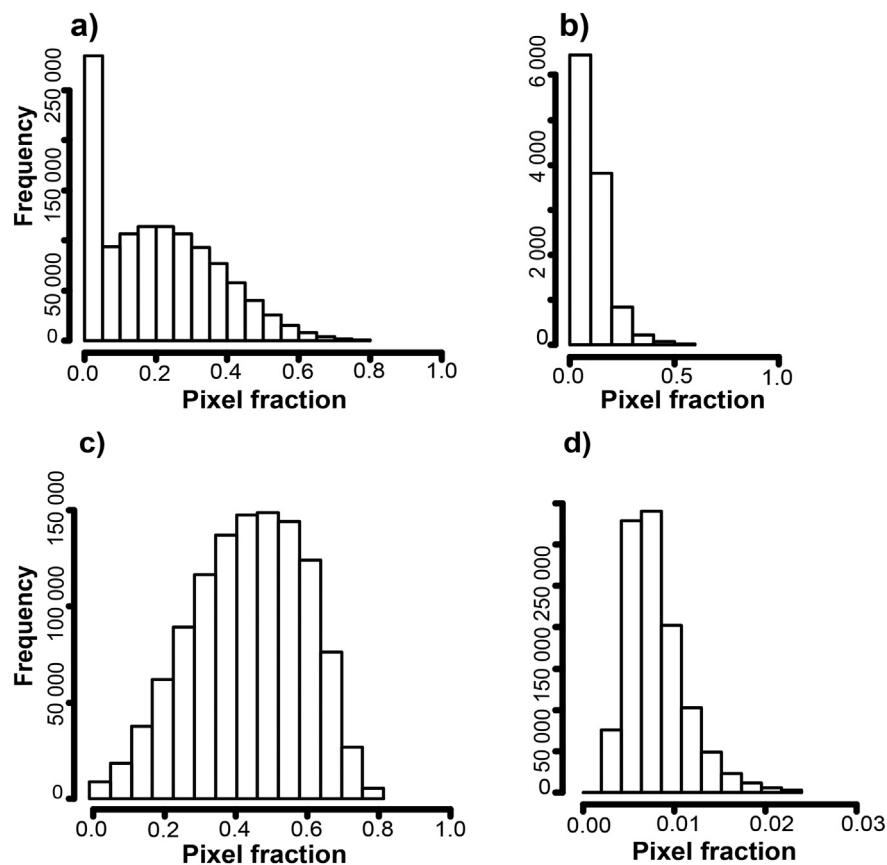


Fig. 6. Frequency of the pixel fractions of the probability of occurrence values within the endmember fraction images of a) Striga, b) soil, c) crops and other weeds, and d) distribution of the root mean square error (RMSE) pixel values.

4.5. Accuracy assessment of Striga detection using MESMA

The Striga classification accuracy assessment focused primarily on two classes of interest i.e. Striga infested and non-infested crop fields. The predicted occurrence of Striga was in a good agreement with the occurrence reference points with an OA of 78%, PA of 79% (error of omission of 21%), and UA of 89% (error of commission of 11%) as shown in Table 3. An F-score of 0.84 for mapping the Striga infestation showed that the MESMA algorithm successfully predicted the occurrence of Striga and its relative abundance within each pixel (Table 3). The PA and UA for the non-Striga fields were lower than those obtained for the Striga-infested fields.

5. Discussion

This study leveraged on the strength and versatility of RF and the robust spatiotemporal resolution of S2 imagery in the cloud-based GEE for cropland mapping. Five spectral indices i.e. NDVI, EVI, MSAVI, the red-edge NDVI, and the red-edge EVI, and vegetation red-edge bands (bands 5, 6, 7 and 8a) were used for the cropland mapping. Thus, adding the indices, red-edge bands and the spatial resolution of the S2 sensor permitted the successful delineation of croplands with high precision, including where the agricultural landscapes are very heterogeneous, fragmented and the fields are small (< 0.1 ha). The use of these indices ensured stable and meaningful comparisons of seasonal and inter-annual changes in vegetation growth and activity thus achieving satisfactory classification results (Belgiu and Csillik, 2018). Other studies have also reported the added-value of the improved S2 spectral and spatial resolution for estimating leaf area index (Sibanda et al., 2019),

mapping LULC (Forkuor et al., 2018), Striga weed (Mudereri et al., 2019a), and in cropland mapping (Xiong et al., 2017a).

Many recent studies have reported very high accuracies in cropland mapping using S2 imagery (Forkuor et al., 2018; Noi and Kappas, 2018). Belgiu and Csillik (2018) reported a 96.19% OA, using S2 in cropland mapping, while Sibanda et al. (2015) confirmed high agreements in quantifying above-ground biomass, using S2 data. The differences between the cropland mapping accuracies observed in the present study and those obtained from other studies could have emanated from the heterogeneous nature of the cropping system and differences in the weeding regimes by the diverse small-holder farmers. This heterogeneity results in multiple spectra and texture captured by the sensors. Additionally, most of the farmers in Rongo practice mixed cropping within small field plots (< 0.1 ha) resulting in the diversity of field sizes, orientation, and shapes. These factors could have caused the 12% error (OA = 88%) margin realized in the cropland classification approach used in this study.

Thus, the UA (commission errors) were significantly lower than the PA (omission errors) for cropland mapping. This was mainly because when training the RF algorithm, we aimed to capture many croplands, thus ensuring low omission errors for the cropland class. Thus, some non-croplands were included as croplands, resulting in higher commission errors for the cropland class. In principle, an algorithm must balance PA and UA (Belgiu and Csillik, 2018; Oliphant et al., 2019). This method is similar to the approach used by Oliphant et al. (2019) who also championed reduction of the error of commission by optimizing the RF algorithm, intending to capture as much cropland area as possible. Similarly, this study aimed to collectively capture all croplands including fallow croplands to holistically test on the occurrence of the devastating Striga weed. The target to reduce the propagation of the

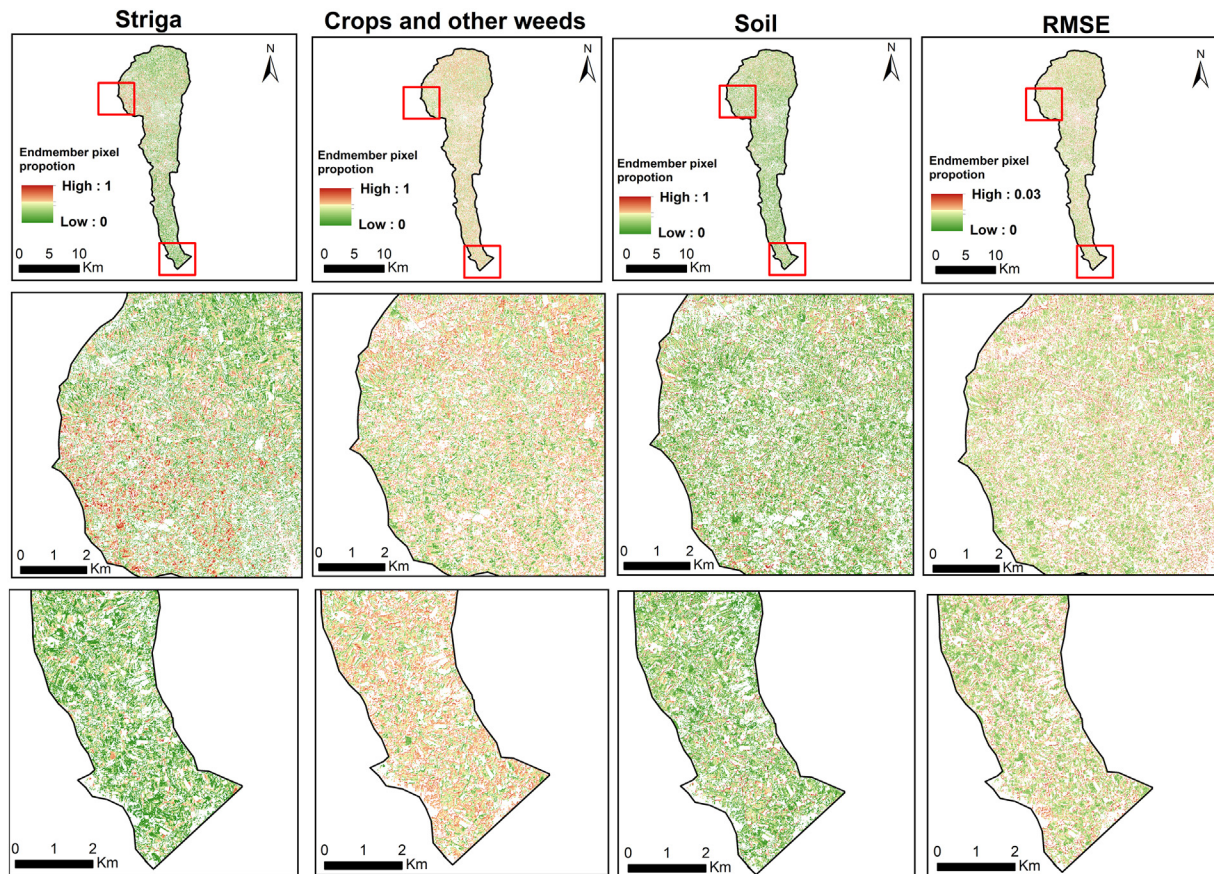


Fig. 7. Classification results from the MESMA showing the fraction images of the three endmembers from the lowest 0 (green) to the highest proportion of 1 (red) and the root mean square error (RMSE) from lowest 0 (green) to the highest 0.03 (red). The images in the four columns represent from left to right: Striga (extreme left column), crops and other weeds, soil, and the respective RMSE (extreme right column). The zoomed-out images show example regions of high Striga proportions (second row from the top) and very low Striga proportions (third row from the top).

error to the MESMA process was achieved, thus the 4% (PA = 96%) cropland omission error is plausible given the damage caused by Striga.

The results produced from the accuracy assessment of the MESMA were assuring, considering that the validation reference data used was entirely independent of the MESMA model, unlike in the standard accuracy assessment procedure where the model internally splits the reference data into training and testing samples. The PA produced using an independent point dataset was relatively high showing relatively low errors of omission in the classification. Other studies have also reported high accuracies when using EMs derived using field spectroradiometer dataset (Landmann, 2003).

Additionally, the novel two-step hierarchical approach for cropland mapping using RF and Striga detection using MESMA improved classification accuracies as the suppression of the spatial constraints (noise) from other classes minimized the spectral confusion (Degerickx et al., 2019). Therefore, applying a few EMs to a spatially constrained agroecosystem reduced the spectral densities and produced relatively high mapping accuracies at low complexity levels. Hence, this hierarchical MESMA approach accounts for the spectral pixel-to-pixel variability of agroecological systems through the simultaneous control of the spatial classification dimension (Franke et al., 2009). Additionally, in contrast with the other classification algorithms, MESMA provides the physical measurement of material contributions in a pixel or vegetation abundance (Li et al., 2005; Rosso et al., 2005). Thus, besides providing information about the occurrence of Striga in Rongo, this study quantified the magnitude and severity of the Striga infestation at suitable spatial resolutions (10 m × 10 m) that have never been provided before. Further studies are nevertheless desirable, to evaluate the performance

of other classifiers and compare them to the Striga detection results derived from the approach of the present study.

The results of the MESMA for detecting Striga revealed that most of the crop fields in Rongo were infested with the Striga weed. The results obtained in this study concurred with the reports from other studies that reported the occurrence of Striga in western Kenya (Atera et al., 2013; Oswald et al., 2001) and in Rongo, mainly in the north and north-western sites of Rongo. This information is critical for agricultural planning in Rongo as most of the farms in the northwestern sites of Rongo practice little crop rotation and concentrate on maize throughout the year on degraded soils. Therefore, there is a need for awareness among the farmers to initiate crop rotation and to improve soil fertility by incorporating cover crops that add organic matter, use green manure, or by growing legumes that enhance nitrogen fixation in the soil. This is particularly different from the southern region where most of the farms are sugarcane plantations, hence the low levels of Striga infestation observed.

Although the use of the three EMs yielded insightful information on Striga weed among co-occurring plants within a heterogeneous agroecological system, the findings should be treated with caution. The location of optimal EMs and their variability to represent the heterogeneity of croplands and their various growth stages over the entire image or in time remains a challenge (Franke et al., 2009). We observed that spectral data from the same class had varying spectral characteristics perhaps attributable to variations in crop ages, insect pests, and disease damage, varying management practices, cropping system, or atmospheric effects (Degerickx et al., 2019). Also, this could be because the pixel size of most satellite-based imaging sensors is slightly bigger

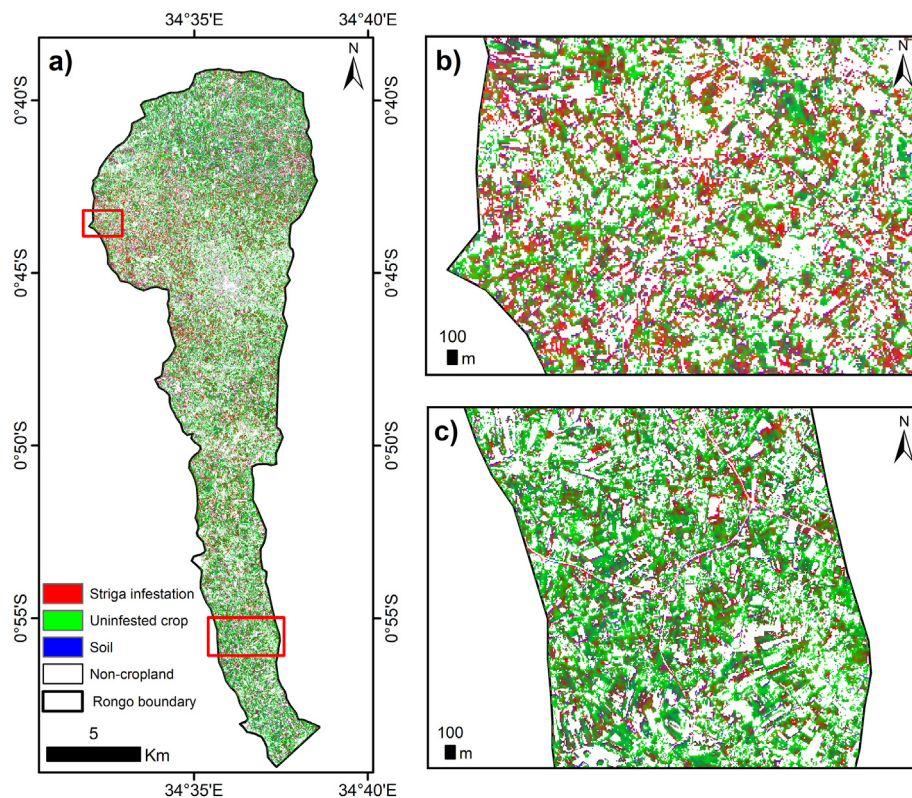


Fig. 8. A graphical representation of the fractions of the three endmembers i.e. Striga, soil and crops and other weeds that were derived from MESMA with a) showing the entire study area, b) zoomed section of the study area with the top red box, and c) zoomed-out section of the study area with the bottom red box.

Table 3

Classification confusion matrix for the Striga and non-Striga classes detected in the Rongo sub-county using Sentinel-2 data and MESMA.

Reference data		Cropland	Non-Striga	Total	UA (%)
Map data	Striga	41	5	46	89
	Non-Striga	11	15	26	58
	Total	52	20	71	
	PA (%)	79	75		
	F-score	0.84	0.65		
OA = 78%					

than the area sensed by the ground-based sensors. Thus, the applied 0.19 m² ASD sample area might have masked out some of the spectral variations in our EMs. This often leads to uncertainties, particularly in consideration of the temporal variation of the EMs spectra captured across different points in time and space.

Therefore, it can be hypothesized that a spectrum that has been tested to adequately model other spectra within the library might not perform with the same precision when applied on a different satellite image or in a different timescale (Song, 2005). These disparities are often introduced by fluxes in the brightness caused by bidirectional reflectance that is exacerbated by large viewing angles and broadband wavelengths when in-situ EMs are upscaled to satellite-based imagery resolutions (Franke et al., 2009; Rosso et al., 2005). However, for this particular study, the brightness-effects did not appear to have had any major impact on the Striga detection results, since MESMA automatically integrates shade as an endmember. Therefore, the use of this approach in other spatial or temporal settings should be applied with caution. Also, this study was based on data collected during one of the Striga peak flowering periods (December), hence there is a need for further research integrating all Striga peak seasons in Kenya (i.e. June and December).

6. Conclusions

In this study, a two-step hierarchical approach was employed for mapping cropland using the RF classifier in GEE and the Striga weed using MESMA on S2 data within a heterogeneous agroecological system in the Rongo sub-county in Kenya. This study confirmed the effectiveness of the GEE as a data curation and cropland characterization platform using RF. The high potential of the MESMA algorithm to decompose mixed pixels and detect Striga occurrence and infestation levels were demonstrated. Therefore, the masking out of other non-target classes for different land cover classes before the implementation of MESMA allowed for a more focused and spatially adjusted spectral unmixing procedure that proved to enhance the classification accuracy results and reduced spectral confusion. Besides, deriving EMs from the in-situ spectrometric dataset provided a more realistic array spectrum of Striga rather than deriving the EMs from the S2 image itself. Future studies should compare the use of Striga EMs from other sources and the respective EM selection mechanisms such as the endmember average root mean squared error (EAR), the minimum average spectral angle (MASA), and the count based endmember selection (CoB) that have been tested in other studies and are known to provide a filter to get pure spectra for use as EMs. Furthermore, Striga infestation fraction estimates could be empirically tested in areas of low, moderate, and high infestation, using the MESMA and cropland mapping hierarchical approach performed in this study.

CRediT authorship contribution statement

All co-authors provided valuable input regarding fieldwork, data analysis, and manuscript preparation. Conceptualization, BTM; EMA; TD and TL; Formal analysis, BTM; Funding acquisition, SN; ZK and TL; Methodology, BTM; Project administration, SN; ZK, and TL; Software, BTM and HEZT; Supervision, EMA; TD; SN and TL; Validation, BTM;

Writing – original draft, BTM.; Writing – review & editing, EMA; TD; SN; ZK; HEZT and TL.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We gratefully acknowledge the financial support for this research by the following organizations and agencies: Biovision Foundation for Ecological Development (Switzerland), grant number (BV DPP-010/2019); UK's Foreign, Commonwealth & Development Office (FCDO); Swedish International Development Cooperation Agency (Sida); the Swiss Agency for Development and Cooperation (SDC); Ethiopian and Kenyan Governments. "B.T.M" 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. *The views expressed herein do not necessarily reflect the official opinion of the donors.*

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>.
- ASD, . FieldSpec® HandHeld 2 User Manual<http://www.geo-informatie.nl/courses/grs60312/material2017/manuals/600860-dHH2Manual.pdf>.
- Asner, G.P., Lobell, D.B., 2000. A biogeophysical approach for automated SWIR unmixing of soils and vegetation. *Remote Sens. Environ.* 74, 99–112. [https://doi.org/10.1016/S0034-4257\(00\)00126-7](https://doi.org/10.1016/S0034-4257(00)00126-7).
- Atera, E.A., Ishii, T., Onyango, J.C., Itoh, K., Azuma, T., 2013. Striga Infestation in Kenya : Status , Distribution and Management Options. vol. 2, pp. 99–108. <https://doi.org/10.5539/sar.v2n2p99>.
- Bannari, A., Morin, D., Bonn, F., Huete, A.R., 1995. A review of vegetation indices. *Remote Sens. Rev.* 13, 95–120. <https://doi.org/10.1080/02757259059532298>.
- Belgiu, M., Csillik, O., 2018. Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. *Remote Sens. Environ.* 204, 509–523. <https://doi.org/10.1016/j.rse.2017.10.005>.
- Bey, A., Jetimane, J., Lisboa, S.N., Ribeiro, N., Siteo, A., Meyfroidt, P., 2020. Mapping small-holder and large-scale cropland dynamics with a flexible classification system and pixel-based composites in an emerging frontier of Mozambique. *Remote Sens. Environ.* 239, 111611. <https://doi.org/10.1016/j.rse.2019.11.1611>.
- Breiman, L., 2001. Randomforest2001 1–33. doi:<https://link.springer.com/content/pdf/10.1023%2FA%3A1010933404324.pdf>.
- 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>.
- Correa Pabón, R.E., Filho, Souza, de, C.R., Oliveira, W.J. de, 2019. Reflectance and imaging spectroscopy applied to detection of petroleum hydrocarbon pollution in bare soils. *Sci. Total Environ.* 649, 1224–1236. <https://doi.org/10.1016/j.scitotenv.2018.08.231>.
- Degerickx, J., Roberts, D.A., Somers, B., 2019. Enhancing the performance of multiple Endmember spectral mixture analysis (MESMA) for urban land cover mapping using airborne lidar data and band selection. *Remote Sens. Environ.* 221, 260–273. <https://doi.org/10.1016/j.rse.2018.11.026>.
- Dube, T., Pandit, S., Shoko, C., Ramoelo, A., Mazvimavi, D., Dalu, T., Dube, T., Pandit, S., Shoko, C., Ramoelo, A., Mazvimavi, D., Dalu, T., 2019. Numerical assessments of leaf area index in tropical Savanna rangelands, South Africa using Landsat 8 OLI derived metrics and in-situ measurements. *Remote Sens.* 11, 829. <https://doi.org/10.3390/rs11070829>.
- Ejeta, G., Gressel, J., 2007. Integrating New Technologies for Striga Control: Towards Ending the Witch-Hunt. World Scientific, Singapore <https://doi.org/10.1142/6470>.
- 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.
- Forkuor, G., Dimobe, K., Serme, I., Tondoh, J.E., 2018. Landsat-8 vs. Sentinel-2: examining the added value of Sentinel-2's red-edge bands to land-use and land-cover mapping in Burkina Faso. *GIScience and Remote Sensing* 55, 331–354. <https://doi.org/10.1080/15481603.2017.1370169>.
- Franke, J., Roberts, D.A., Halligan, K., Menz, G., 2009. Hierarchical multiple endmember spectral mixture analysis (MESMA) of hyperspectral imagery for urban environments. *Remote Sens. Environ.* 113, 1712–1723. <https://doi.org/10.1016/j.rse.2009.03.018>.
- Graesser, J., Ramankutty, N., 2017. Detection of cropland field parcels from Landsat imagery. *Remote Sens. Environ.* 201, 165–180. <https://doi.org/10.1016/j.rse.2017.08.027>.
- Gumma, M.K., Thenkabail, P.S., Teluguntla, P.G., Oliphant, A., Xiong, J., Giri, C., Pyla, V., Dixit, S., Whitbread, A.M., 2019. Agricultural cropland extent and areas of South Asia derived using Landsat satellite 30-m time-series big-data using random forest machine learning algorithms on the Google earth engine cloud. *GIScience and Remote Sensing* 00, 1–21. <https://doi.org/10.1080/15481603.2019.1690780>.
- Hassanali, A., Herren, H., Khan, Z.R., Pickett, J.A., Woodcock, C.M., 2008. Integrated pest management: the push-pull approach for controlling insect pests and weeds of cereals, and its potential for other agricultural systems including animal husbandry. *Philosophical Transactions of the Royal Society B: Biological Sciences* 363, 611–621. <https://doi.org/10.1098/rstb.2007.2173>.
- Heinz, D.C., Chang, C.I., 2001. Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* 39, 529–545. <https://doi.org/10.1109/36.911111>.
- Huete, A., Didan, K., Miura, T., Rodriguez, E., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* 83, 195–213. [https://doi.org/10.1016/S0034-4257\(02\)00096-2](https://doi.org/10.1016/S0034-4257(02)00096-2).
- Huovinen, P., Ramírez, J., Palacios, M., Gómez, I., 2020. Satellite-derived mapping of kelp distribution and water optics in the glacier impacted Yendegai Fjord (Beagle Channel, southern Chilean Patagonia). *Sci. Total Environ.* 703, 135531. <https://doi.org/10.1016/j.scitotenv.2019.135531>.
- 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>.
- Kelley, L.C., Pitcher, L., Bacon, C., 2018. Using Google Earth Engine to map complex shade-grown coffee landscapes in northern Nicaragua. *Remote Sens.* 10, 952. <https://doi.org/10.3390/rs10060952>.
- 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>.
- Landmann, T., 2003. Characterizing sub-pixel Landsat ETM+ fire severity on experimental fires in the Kruger National Park, South Africa. *S. Afr. J. Sci.* 99, 357–360.
- Landmann, T., Eidmann, D., Cornish, N., Franke, J., Siebert, S., 2019. Optimizing harmonics from Landsat time series data: the case of mapping rainfed and irrigated agriculture in Zimbabwe. *Remote Sensing Letters* 10, 1038–1046. <https://doi.org/10.1080/2150704X.2019.1648901>.
- Leutner, B., Horning, N., Schwalb-willmann, J., 2019. Tools for Remote Sensing Data Analysis. Rpackage Version 0.2.4.
- Li, L., Ustin, S.L., Lay, M., 2005. Application of multiple endmember spectral mixture analysis (MESMA) to AVIRIS imagery for coastal salt marsh mapping: a case study in China camp, CA, USA. *Int. J. Remote Sens.* 26, 5193–5207. <https://doi.org/10.1080/0143160500218911>.
- Lukas, W., Lehnert, Hanna, M., Joerg, B., 2018. hsdr: Manage, Analyse and Simulate Hyperspectral Data in R. R Package Version 0.7.2.
- Mudereri, B.T., Dube, T., Adel-Rahman, E.M., Niassy, S., Kimathi, E., Khan, Z., Landmann, T., 2019a. A comparative analysis of PlanetScope and Sentinel-2 space-borne sensors in mapping Striga weed using guided regularised random forest classification ensemble. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-2/W13*, 701–708. <https://doi.org/10.5194/isprs-archives-XLII-2-W13-701-2019>.
- Mudereri, B.T., Chitata, T., Mukanga, C., Mupfiga, E.T., Gwatisira, C., Dube, T., 2019b. Can biophysical parameters derived from Sentinel-2 spaceborne sensor improve land cover characterization in semi-arid regions? *Geocart International* <https://doi.org/10.1080/10106049.2019.1695956>.
- Mudereri, B.T., Dube, T., Niassy, S., Kimathi, E., Landmann, T., Khan, Z., Abdel-Rahman, E.M., 2020. Is it possible to discern Striga weed (*Striga hermonthica*) infestation levels in maize agro-ecological systems using in-situ spectroscopy? *Int. J. Appl. Earth Obs. Geoinf.* 85, 102008. <https://doi.org/10.1016/j.jag.2019.102008>.
- Mupfiga, E.T., Munkwakwata, R., Mudereri, B.T., Nyatondo, U.N., 2016. Assessment of sedimentation in Tuli – Makwe dam using remotely sensed data. *Journal of Soil Science and Environment and Management* 4, 127–136.
- Mutanga, O., Dube, T., Galal, O., 2017. Remote sensing of crop health for food security in Africa: potentials and constraints. *Remote Sensing Applications: Society and Environment* 8, 231–239. <https://doi.org/10.1016/j.rsase.2017.10.004>.
- Njenga, G.W., 2016. Multiple Endmember Spectral Mixture Analysis (MESMA) on Multi-temporal VHR Images for Weed Detection in Smallholder Farms. p. 116.
- Noi, P.T., Kappas, M., 2018. Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery. *Sensors (Switzerland)*, 18 <https://doi.org/10.3390/s18010018>.
- Oliphant, A.J., Thenkabail, P.S., Teluguntla, P., Xiong, J., Krishna, M., Congalton, R.G., Yadav, K., 2019. Mapping cropland extent of Southeast and Northeast Asia using multi-year time-series Landsat 30-m data using a random forest classifier on the Google Earth Engine Cloud. *Int. J. Appl. Earth Obs. Geoinf.* 81, 110–124. <https://doi.org/10.1016/j.jag.2018.11.014>.
- Oswald, A., Ransom, J.K., Kroschel, J., Sauerborn, J., 2001. Transplanting maize and sorghum reduces *Striga hermonthica* damage. *Weed Sci.* 49, 346–353. [https://doi.org/10.1614/0043-1745\(2001\)049\[0346:TMASRS\]2.0.CO;2](https://doi.org/10.1614/0043-1745(2001)049[0346:TMASRS]2.0.CO;2).
- Powell, R.L., Roberts, D.A., Dennison, P.E., Hess, L.L., 2007. Sub-Pixel Mapping of Urban Land Cover Using Multiple Endmember Spectral Mixture Analysis: Manaus , Brazil. vol. 106, pp. 253–267. <https://doi.org/10.1016/j.rse.2006.09.005>.

- Qi, J., Chehbouni, A., Huete, A.R., Kerr, Y.H., Sorooshian, S., 1994. A modified soil adjusted vegetation index. *Remote Sens. Environ.* 48, 119–126. [https://doi.org/10.1016/0034-4257\(94\)90134-1](https://doi.org/10.1016/0034-4257(94)90134-1).
- Quintano, C., Fernández-Manso, A., Roberts, D.A., 2013. Multiple Endmember spectral mixture analysis (MESMA) to map burn severity levels from Landsat images in Mediterranean countries. *Remote Sens. Environ.* 136, 76–88. <https://doi.org/10.1016/j.rse.2013.04.017>.
- R Core Team, 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Roberts, D.A., Gardner, M., Church, R., Ustin, S., Scheer, G., Green, R.O., 1998. Mapping chaparral in the Santa Monica Mountains using multiple endmember spectral mixture models. *Remote Sens. Environ.* 65, 267–279. [https://doi.org/10.1016/S0034-4257\(98\)00037-6](https://doi.org/10.1016/S0034-4257(98)00037-6).
- Rosso, P.H., Ustin, S.L., Hastings, A., 2005. Mapping marshland vegetation of San Francisco Bay, California, using hyperspectral data. *Int. J. Remote Sens.* 26, 5169–5191. <https://doi.org/10.1080/01431160500218770>.
- Rouse, J.W., Hass, R.H., Schell, J.A., Deering, D.W., 1974. Monitoring vegetation systems in the great plains with ERTS. *Third Earth Resources Technology Satellite (ERTS) Symposium 1*, pp. 309–317 (doi:citeulike-article-id:12009708).
- 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>.
- Sibanda, M., Mutanga, O., Dube, T., S Vundla, T., L Mafongoya, P., 2019. Estimating LAI and mapping canopy storage capacity for hydrological applications in wattle infested ecosystems using Sentinel-2 MSI derived red edge bands. *GIScience and Remote Sensing* 56, 68–86. <https://doi.org/10.1080/15481603.2018.1492213>.
- Somers, B., Asner, G.P., Tits, L., Coppin, P., 2011. Endmember variability in spectral mixture analysis : a review. *Remote Sens. Environ.* 115, 1603–1616. <https://doi.org/10.1016/j.rse.2011.03.003>.
- Song, C., 2005. Spectral mixture analysis for subpixel vegetation fractions in the urban environment: how to incorporate endmember variability? *Remote Sens. Environ.* 95, 248–263. <https://doi.org/10.1016/j.rse.2005.01.002>.
- Thenkabail, P.S., Lyon, J.G., Huete, A., 2019. Biophysical and biochemical characterization and plant species studies. *Hyperspectral Remote Sensing of Vegetation*, Second Edi CRC Press, Boca Raton <https://doi.org/10.1201/9780429431180>.
- Wang, J.J., Zhang, Y., Bussink, C., 2014. Unsupervised multiple endmember spectral mixture analysis-based detection of opium poppy fields from an EO-1 Hyperion image in Helmand, Afghanistan. *Sci. Total Environ.* 476–477, 1–6. <https://doi.org/10.1016/j.scitotenv.2014.01.006>.
- Wu, W., Luo, J., Hu, X., Yang, H., Yang, Y., 2018. A thin-cloud mask method for remote sensing images based on sparse dark pixel region detection. *Remote Sens.* 10. <https://doi.org/10.3390/rs10040617>.
- Xiong, J., Thenkabail, P.S., Gumma, M.K., Teluguntla, P., Poehnelt, J., Congalton, R.G., Yadav, K., Thau, D., 2017a. Automated cropland mapping of continental Africa using Google Earth Engine cloud computing. *ISPRS J. Photogramm. Remote Sens.* 126, 225–244. <https://doi.org/10.1016/j.isprsjprs.2017.01.019>.
- Xiong, J., Thenkabail, P.S., Tilton, J.C., Gumma, M.K., Teluguntla, P., Oliphant, A., Congalton, R.G., Yadav, K., Gorelick, N., 2017b. Nominal 30-m cropland extent map of continental Africa by integrating pixel-based and object-based algorithms using Sentinel-2 and Landsat-8 data on google earth engine. *Remote Sens.* 9, 1–27. <https://doi.org/10.3390/rs9101065>.
- Xue, J., Su, B., 2017. Significant remote sensing vegetation indices: a review of developments and applications. *Journal of Sensors* 2017. <https://doi.org/10.1155/2017/1353691>.