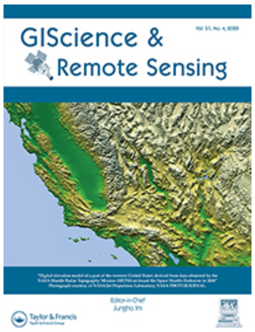
GIScience & Remote Sensing

ISSN: 1548-1603 (Print) 1943-7226 (Online) Journal homepage: https://www.tandfonline.com/loi/tgrs20



Multi-source spatial data-based invasion risk modeling of Striga (*Striga asiatica*) in Zimbabwe

Bester Tawona Mudereri, Elfatih Mohamed Abdel-Rahman, Timothy Dube, Tobias Landmann, Zeyaur Khan, Emily Kimathi, Rachel Owino & Saliou Niassy

GISCIENCE & REMOTE SENSING
2020, VOL. 57, NO. 4, 553–571
https://doi.org/10.1080/15481603.2020.1744250

Taylor & Francis
Taylor & Francis Group

Multi-source spatial data-based invasion risk modeling of Striga (*Striga asiatica*) in Zimbabwe

Bester Tawona Mudereri[a,b], Elfatih Mohamed Abdel-Rahman[a,c], Timothy Dube[b], Tobias Landmann[a,d], Zeyaur Khan[a], Emily Kimathi[a], Rachel Owino[a] and Saliou Niassy[a]

[a]International Centre of Insect Physiology and Ecology (icipe), Nairobi, Kenya; [b]Department of Earth Sciences, University of Western Cape, Bellville, South Africa; [c]Department of Agronomy, Faculty of Agriculture, University of Khartoum, Khartoum, North Sudan; [d]RSS-Remote Sensing Solutions Gmbh, Munich, Germany

ABSTRACT

Monitoring of destructive invasive weeds such as those from the genus *Striga* requires accurate, near real-time predictions and integrated assessment techniques to enable better surveillance and consistent assessment initiatives. Thus, in this study, we predicted the potential ecological niche of Striga (*Striga asiatica*) weed in Zimbabwe, to identify and understand its propagation and map potentially vulnerable cropping areas. Vegetation phenology from remote sensing, bioclimatic and other environmental variables (i.e. cropping system, edaphic, land surface temperature, and terrain) were used as predictors. Six machine learning modeling techniques and the ensemble model were evaluated on their suitability to predict current and future Striga weed distributional patterns. The mentioned predictors (n = 40) were integrated into six models with "presence-only" training and evaluation data, collected in Zimbabwe over the period between the 12th and 28th of March 2018. The area under the curve (AUC) and true skill statistic (TSS) were used to measure the performance of the Striga modeling framework. The results showed that the ensemble model had the strongest Striga occurrence predictive power (AUC = 0.98; TSS = 0.93) when compared to the other modeling algorithms. Temperature seasonality (Bio4), the maximum temperature of the warmest month (Bio5) and precipitation seasonality (Bio15) were determined to be the most dominant bioclimatic variables influencing Striga occurrence. "Start of the season" and "season minimum value" of the "Enhanced Vegetation Index base value" were the most relevant remote sensing-based variables. Based on projected climate change scenarios, the study showed that up to 2050, the suitable area for Striga propagation will increase by ~ 0.73% in Zimbabwe. The present work demonstrated the importance of integrating multi-source data in predicting possible crop production restraints due to weed propagation. The results can enhance national preparedness and management strategies, specifically, if the current and future risk areas can be identified for early intervention and containment

ARTICLE HISTORY
Received 10 October 2019
Accepted 13 March 2020

KEYWORDS
Climate variability; food security; machine learning; niche modeling; remote sensing; sub-Saharan Africa; Striga weeds

Introduction

The genus *Striga* is composed of several species of parasitic weeds of global economic importance that cause socioeconomic losses of over 1 billion USD in sub-Saharan Africa (Spallek, Musembi, and Shirasu 2013). *Striga hermonthica* and *S. asiatica* are the most prevalent among the Striga species with *S. asiatica* reported affecting approximately 40% of arable land in the region (Cochrane and Press 1997). These two species thrive in climatic conditions, which are also favorable for most cereal crops like maize, sorghum, millet and rice (Khan et al. 2014). These cereal crops are the most important staple crops for the majority of the African population, with maize being grown for food purposes by more than 300 million people out of an estimated one billion population in sub-Saharan Africa (SSA: Sasson 2012).

Striga weeds attach themselves to the roots of the cereal crops, after germination and outcompete their hosts for space, nutrients, and water. As a result, the growth and development of the host crops deteriorate, causing considerable yield reduction (Ejeta and Gressel 2007). In some areas in Africa, the scourge of Striga has reached epidemic magnitudes, causing a desperate scenario, mostly to poor small-scale farmers (Mandumbu et al. 2017b). The most common response practice in such scenarios is for farmers to abandon the land and look for new croplands, a very labor-intensive task that inevitably contributes to cropland expansion and severe environmental degradation.

In the present study, we opted to predict the probability of occurrence of Striga (i.e. *S. asiatica*) in Zimbabwe, using vegetation phenology from remote sensing, bioclimatic, other remotely sensed variables (i.e. cropping system, edaphic, land surface temperature, and terrain), empirical

CONTACT Bester Tawona Mudereri bmudereri@icipe.org

machine learning (ML) and ecological niche modeling (ENM) approaches. The advent of these freely available earth observation "big data" from multiple sources and ML algorithms permit access to a new paradigm of immense opportunities to understand the earth and agroecological systems over time and space (Cian, Marconcini, and Ceccato 2018). This allows for comprehensive statistical analysis on large temporal resolution data using the phenological characteristics hidden in these time-series data (Landmann et al. 2019). These various time-series inherent characteristics in the "big data," are often concealed in single snapshot remotely sensed imagery (Cian, Marconcini, and Ceccato 2018; Ochungo et al. 2019).

Thus, these multi-source remotely sensed data, coupled with advanced and efficient ML and ENM approaches provide a cost-effective, timely, robust and very accurate platform to map and predict the occurrence of invasive weeds like Striga (Thamaga and Dube 2019; Jafarian, Kargar, and Bahreini 2019). In particular, mapping flowering Striga, i.e. *S. hermonthica* using in-situ methods, high spatial resolution satellite data and ML has been proven to be largely possible and achievable at the plot (Mudereri et al. 2020a) and field scale (Mudereri et al. 2019a). However, the potential of using these remotely sensed data to detect and map the risk posed by the understory Striga such as the *S. asiatica*, which exists completely covered underneath crop canopies has not been attempted anywhere, more so at landscape scales or by using multi-source data. This deficit in information is mainly attributable to the heterogeneous nature of the agro-natural landscapes in Africa and the multiple spectral responses obtained from crop fields that are infested with understory Striga weed, which cause enormous errors in their detection and mapping (Mudereri et al. 2019a).

When used in species distribution modeling, ML and ENMs correlate the present location ("presence-only" or "presence-absence" data) of a species with the appropriate predictor variables (e.g. environmental variables), thereby providing a statistical link between the spatial differences of the predictor variables and the dispersion of the species in the environment, in our case Striga (Ayebare et al. 2018). Accuracy of the ML and ENM relies on the precision and distribution of the "presence-only" data tied with a careful selection of ecological and climatic predictor variables (Elith et al. 2010). However, it can be inferred that there is no universally best ML algorithm, which warrants the scoping into the best predictive model and the best predictor combinations for species distribution (Guo et al. 2019).

Therefore, identifying robust ML and ENM modeling algorithms that can select the most relevant predictor variables from multiple ecological covariates to predict the occurrence, propagation and distribution of the understory Striga species such as boosted regression trees (BRT: Friedman 2001), classification and regression trees (CART: Breiman et al. 1984), flexible discriminant analysis (FDA: Fisher 1936), generalized linear model (GLM: Nelder and Wedderburn 1972), random forest (RF: Breiman 2001) and support vector machines (SVM: Vapnik 1979), is crucial. Moreover, integrating the remotely sensed and bioclimatic data in such ML and ENM has been reported in other studies as the best way to produce reliable and accurate results by harnessing the vast information provided by the intrinsic phenological vegetation metrics and the external influence of climatic variables (Kyalo et al. 2018; Makori et al. 2017).

However, there is a huge deficit in information regarding the use of such technology on analysis and mapping of Striga distribution and risk particularly the influence of climate change on the distribution of Striga in Africa. Many studies have investigated the effects of climate change on a range of species, showing that change in climatic conditions has a profound impact on species distribution ranges (Mbatudde et al. 2012; Wan and Wang 2019; Guan et al. 2020). To the best of our knowledge, there is no precise spatial information or scenario modeling detailing the current or probable risk that climate change will impose on the distribution, occurrence, and severity of Striga in Zimbabwe or on the entire African continent. The risk inflicted by these parasitic weeds is likely to be worsened by climate change and the inadequate adaptive or mitigation capacity, in addition to the limited impact documentation leading to inadequate preparedness (Niang et al. 2014).

Thus, identifying and controlling these invasive weeds before they can spread to new environments requires better surveillance and constant monitoring across the African countries with adequate, cost-effective tools and methods . We, therefore, hypothesized that climate change might cause restrictions or expansions on the distribution of Striga species through altering host availability or imposing Striga intolerable or suitable climatic conditions (Cochrane and Press 1997). Thus, the uncertainties brought by these future climate scenarios necessitate robust and accurate mapping methods and relevant environmental multi-source variables and datasets to estimate and predict the potential and actual impact of climate change on the current and future distribution of the biological niche of Striga in Zimbabwe.

Study area

We predicted the occurrence of Striga in the 10 provinces of Zimbabwe (Figure 1). Zimbabwe is

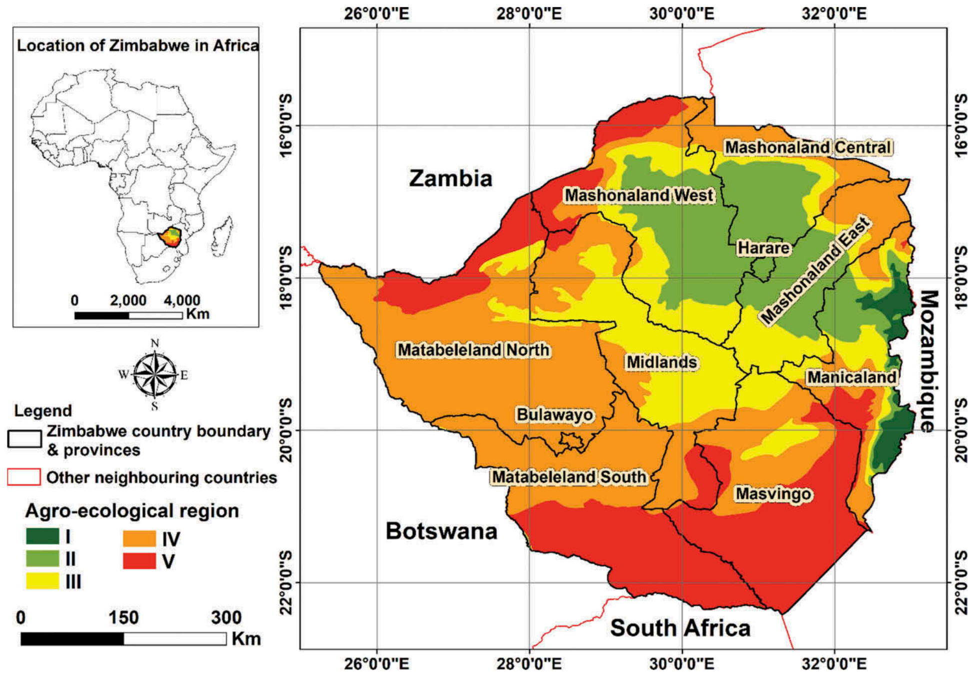


Figure 1. Location of Zimbabwe in Africa and the relative location and boundaries of the five agro-ecological regions of the country which characterize the study area. See Table 1 for a detailed description of the agro-ecological regions.

a landlocked country in southern Africa covering a land area of ~ 390 753 km^2, which shares borders with Botswana, Mozambique, South Africa, Zambia and partly Namibia. It is bound within latitudes 15.6° and 22.4° South and longitudinally between 25.2° and 33.1° East (Kuri et al. 2018). Geographically, the central part of the country is located on a high plateau forming a watershed between the two major river systems, i.e. Zambezi river in the north and Limpopo river on the south. The country has a remarkably varied climate, marked by the differences in latitude which characterizes the wide-ranging rainfall patterns and extensive agronomic activities. It is situated within the tropics and experiences the short cold, dry season between May and September, while the period November to April is marked by heavy rainfall (Mudereri et al. 2019b). Zimbabwe is subdivided into five agro-ecological regions that vary in temperature, rainfall, soil and agricultural potential (Table 1). These five agro-ecological regions include regions I and II referred to as the Highveld; region III which is Middleveld, while region IV and V are referred to as the Lowveld (Sungirai et al. 2018). In Zimbabwe, the lowest annual rainfall is 400 mm which is received in region V and the highest amount (1 200 mm) is received in region I. The mean annual temperature ranges from 16°C in the regions I and II to ~ 26–35°C in the southern Lowveld (Kuri et al. 2019). Approximately, 11% of the country is arable land with ~ 0.31% of that arable land being continuously under different crops such as maize, wheat, sorghum,

Table 1. Characteristics of the five agro-ecological regions of Zimbabwe (Mugandani et al. 2012; FAO, and ACFD 1999).

Agro-ecological region	Average annual rainfall (mm year^{-1})	Mean maximum temperature (^{0}C)	Dominant soil type	Agriculture potential and farming system
I	>1000	16–19	Acrisols, Ferralsols	Suitable for dairy farming forestry, tea, coffee, fruit, beef, and maize production
II	700–1050	19–23	Cambisols, Luvisols, Arenosols	Suitable for intensive farming, based on maize, tobacco, cotton, and livestock
III	500–800	23–26	Arenosols	Suitable for intensive farming, based on maize, tobacco, cotton, and livestock
IV	450–650	19–26	Leptosols, Vertisols, Solonetz	Semi-extensive region. Suitable for farm systems based on livestock and resistant fodder crops. Forestry, wildlife/tourism
V	<450	26–32	Leptosols, Vertisols, Solonetz	Extensive farming region. Suitable for extensive cattle ranching. Zambezi Valley is infested with tsetse fly. Forestry, wildlife/tourism

and millet. Of these crops, maize is the most economically important and commonly grown cereal crop at both small- and large-scales (Kuri et al. 2018).

Methodology

Figure 2 shows a flowchart that explains the entire methodology adopted in the present study for modeling Striga invasion risk.

Striga occurrence data collection

The Striga "presence-only" data were collected between the period 12th and 28th of March 2018. The data collection period matched with the flowering phase of Striga in Zimbabwe. We targeted the flowering phase since this is the best time to differentiate Striga from other co-occurring weeds (Mudereri et al. 2020a). Reference "presence-only" data were gathered within maize croplands, which in our study area were mainly mono and mixed maize cropping systems. The mixed cropping system is mostly a combination of maize with groundnuts, round-nuts or beans. We employed a purposive sampling approach using local expert knowledge (i.e. extension officers and farmers) to assist in the identification of the Striga infested fields. A handheld global positioning system (GPS) device with an error margin of ±3 m was used to locate the reference control points. We avoided the edge-effect by collecting the Striga "presence-only" data at the center of the field (sampling unit: 30 m x 50 m). A total of 50 "presence-only" Striga reference data were collected covering the six districts in Midlands and Masvingo provinces (Figure 1); namely Bikita, Chivi, Gweru, Masvingo, Shurugwi and Zaka (agro-ecological regions III and IV). These points were spread across the different elevation gradients (400–1 600 m a.s.l), except for the highest elevation in the Eastern highlands (>1 600 a.s.l). Agro-ecological region V, which was not sampled, is the region mainly reserved for livestock

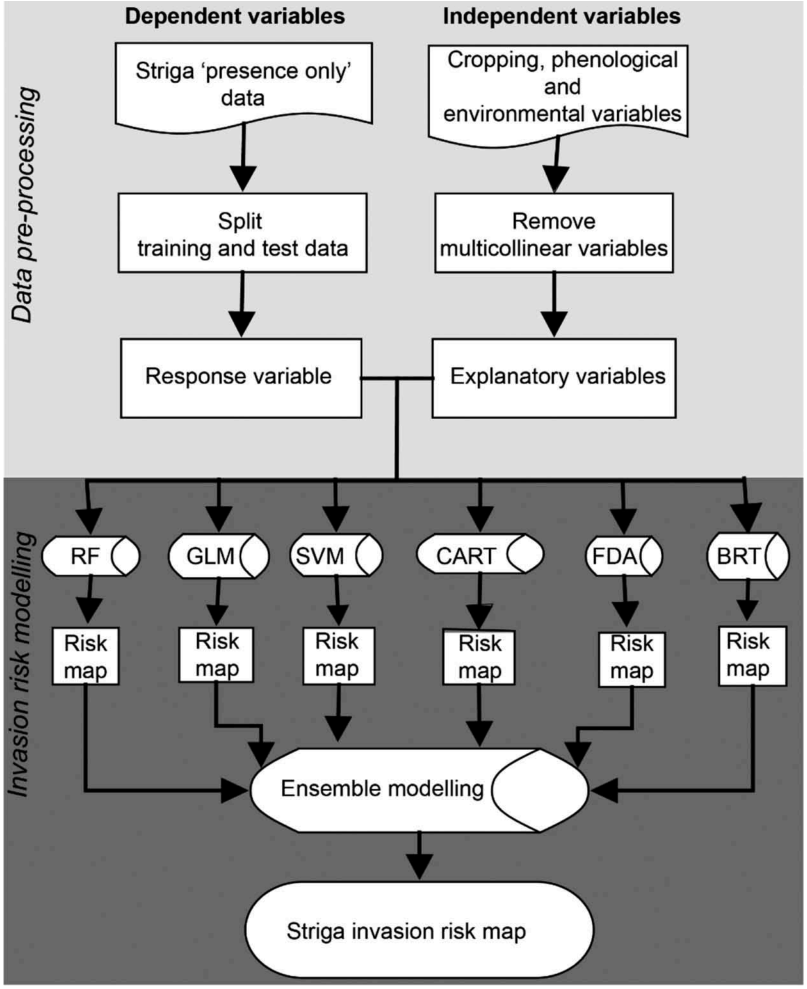


Figure 2. Flow diagram of the methodology of Striga invasion risk modeling. The six models used are random forest (RF), generalized linear model (GLM), support vector machines (SVM), classification and regression trees (CART), flexible discriminant analysis (FDA) and boosted regression trees (BRT).

production in Zimbabwe and is characterized by non-arable land and pastures (*see* Table 1).

Predictor variables

The predictor variables that we used in the present study were grouped into two main categories, i.e. bioclimatic and remotely sensed variables (Tables 2 and 3). Variable spatial and temporal resolutions are a key notion in determining a dataset's fitness for a given use as they influence the pattern that can be observed during the analysis (Degbelo and Kuhn 2018). However, Csillag, Kummert, and Kertész (1992) pointed out that there is no single best resolution when combining environmental variables of varying resolutions. In our case, we had variables ranging in pixel size from 1 km x 1 km to approximately 250 m x 250 m spatial resolution. This variation influences the integration of multiresolution variables within models. We, therefore, counteracted the variation by resampling all the datasets to the lowest spatial resolution of 250 m x 250 m pixel size. In addition, the vegetation phenological variables were derived from multidate input data, while other remotely sensed and the bioclimatic variables were generic. Thus, only their respective derivatives (i.e. output variables) from the TIMESAT model were used as inputs in the invasion risk modeling analysis to offset the effect of the temporal variation. Therefore, we do not anticipate any influence of spatial or temporal resolution on the accuracy of our models. It is further worth noting that all our predictor variables are freely available.

Table 2. Bioclimatic variables used in the species distribution models for Striga occurrence prediction and their variance inflation factor (VIF) values. The variables in bold were used in the final Striga occurrence prediction after eliminating highly correlated ones.

BioClim Code	Environmental variable description	Unit	VIF value
Bio1	**Annual mean temperature**	**^{0}C**	**7.30**
Bio2	Mean diurnal range [mean of monthly (max temp−min temp)]	^{0}C	3.05
Bio3	**Iso-thermality (Bio2/Bio7) (×100)**		**3.25**
Bio4	**Temperature seasonality (standard deviation ×100)**		**2.00**
Bio5	**Maximum temperature of the warmest month**	**^{0}C**	**8.21**
Bio6	Min temperature of the coldest month	^{0}C	10.23
Bio7	Temperature annual range (Bio5–Bio6)	^{0}C	12.63
Bio8	Mean temperature of wettest quarter	^{0}C	5.04
Bio9	Mean temperature of driest quarter	^{0}C	1.92
Bio10	Mean temperature of warmest quarter	^{0}C	8.73
Bio11	Mean temperature of coldest quarter	^{0}C	1.62
Bio12	Annual precipitation	mm	2.76
Bio13	Precipitation of wettest month	mm	7.27
Bio14	Precipitation of driest month	mm	3.91
Bio15	**Precipitation seasonality (coefficient of variation)**	**mm**	**4.32**
Bio16	Precipitation of wettest quarter	mm	4.70
Bio17	Precipitation of driest quarter	mm	6.33
Bio18	**Precipitation of warmest quarter**	**mm**	**5.41**
Bio19	Precipitation of coldest quarter	mm	2.58

Bioclimatic variables

We used 19 bioclimatic variables (Table 2) that are freely downloadable from the WorldClim platform (www.worldclim.org) at ~1 km x 1 km spatial resolution to determine the key climatic conditions influencing the distribution of Striga in Zimbabwe for both the current and future climate scenarios. Four representative concentration pathways (RCPs) were set by the intergovernmental panel on climate change (IPCC) using the total radioactive forcing of values 2.6, 4.5, 6.0 and 8.5 watt/m^2 (IPCC 2014). In this study, we only used the current bioclimatic data (-1950–2000) and a one-time step of the future climate data of the maximum emission (RCP8.5) for the CO_2 concentrations predicted for 2050 (average of predictions for 2041–2060: Guan et al. 2020). The future climatic data were obtained from the fourth version of the community climate system model (CCSM4), which is one of the models that provide the most efficient global future climate projections (Mohammadi et al. 2019; Mudereri et al. 2020b). All these bioclimatic variables were clipped to the Zimbabwean country boundary and resampled to 250 m x 250 m pixel size, to match the size and extents of the remotely sensed variables.

Remotely sensed variables

We used a total of five remotely sensed variable categories: cropping system, edaphic, land surface temperature, terrain and vegetation phenology (vegetation seasonality characteristics) (Table 3). These variables were selected because they were reported in several studies as key determinants of Striga distribution (Mandumbu et al. 2017a; Oswald et al. 2001; Parker 2009; Mudereri et al. 2019a). All the remotely sensed variables that were not in the 250 m x 250 m spatial resolution were standardized and resampled to this pixel size.

Cropping system variable. We used the cropping system variable provided by the study of Landmann et al. (2019). The variable was obtained at 30 m x 30 m pixel size with three categorical classes: rainfed wildland, rainfed cropland, and irrigated cropland. The cropping system variable for Zimbabwe was derived at 97% accuracy (Landmann et al. 2019) using vegetation harmonics of the time-series normalized difference vegetation index (NDVI) derived from Landsat 8 operational land imager (OLI) images.

Table 3. Remotely sensed variables used in the species distribution models for Striga occurrence prediction and their variance inflation factor (VIF) values. The variables in bold were used in the final Striga occurrence prediction model after eliminating highly correlated ones. EVI is the enhanced vegetation index.

Variable	Description	Units	VIF Value
Cropping system			
Cropping system	**Irrigated or rain-fed cropland/wildland**	**Categorical**	**1.20**
Edaphic variables			
Sand content	**Quantity of sand in the soil**	**g/100 g**	**2.44**
Soil nitrogen (N)	**Total amount of nitrogen in the soil**	**mg/kg**	**2.50**
Soil organic carbon	**Organic matter present in the soil**	**g/kg**	**1.89**
Soil pH	**Acidity or alkalinity of the soil**	**pH value**	**2.07**
Land surface temperature (LST)			
LST	**Land surface temperature**	**K**	**1.07**
Seasonality variables			
Amplitude	The difference between the maximum EVI and base value	EVI value	1.09
Base value	**Minimum EVI value**	**EVI value**	**1.60**
End of season	**EVI value at the time of the end of season**	**EVI value**	**1.02**
Large integral	**Integral of the season from start to end**		**1.58**
Left derivative	**Rate of EVI value increase at the start of the season**	**%**	**1.23**
Length of season	Time-lapse from start to end of season	Days	6.88
Maximum EVI	Maximum EVI value in season	EVI value	1.05
Middle of season	Absolute value at the middle of the season	EVI value	2.19
Right derivative	Rate of EVI value increase at the end of season	%	1.99
Small integral	Integral of the season and base value from start to end of season		2.47
Start of season	**EVI value at the beginning of the season**	**EVI value**	**1.04**
Terrain variables			
Aspect	**Slope direction**	**Degrees**	**1.15**
Elevation	**Ground height above sea level**	**m**	**4.46**
Hillshade	**Shading the sun effect**		**1.22**
Slope	**Ground steepness**	**%**	**1.76**

Edaphic variables. We used four soil properties that were downloaded from https://www.isric.org/explore and referred to as the "AfSoilGrids250 m" (Hengl et al. 2015). Specifically, we used the Africa soil grids produced at 0–30 cm depth with a spatial resolution of 250 m x 250 m. The soil characteristics data includes total soil nitrogen (N) (mg/kg: ppm), soil pH, soil organic carbon (SOC) in g/kg and sand content (50–2000 µm) in g/100 g (Hengl et al. 2015). These variables were chosen because they broadly influence soil fertility, and thus the potential occurrence of Striga. Several studies have established that depleted soil fertility leads to rapid propagation and thriving of Striga within croplands (Ekeleme et al. 2014; Yoneyama et al. 2007).

Land surface temperature. We used "day time" Land Surface Temperature Climate Modeling Grid (LST_Day_CMG) available in K, simulated from Moderate Resolution Imaging Spectro-radiometer (MODIS) data and available at https://lpdaac.usgs.gov/products/mod11c2v006/, (Wan, Hook, and Hulley 2015). Specifically, we used the "multi-day" MOD11C2 LST product of 5.6 × 5.6 km spatial resolution available from the year 2000 to the present. We chose LST because the cereal chemical that triggers Striga germination requires optimal temperature (i.e. 22°–30° C) and Striga seeds also need an optimum soil temperature range to germinate (Rich and Ejeta 2007). We, therefore postulated that the surface fluxes measured by LST would be one of the proxy key variables that immensely predict the potential germination of Striga seeds.

Terrain variables. The terrain variables were calculated from the shuttle radar topographic mission (SRTM) data which are available as 3 arc *sec* (~ 90 m resolution) digital elevation model (DEM). The vertical error of our DEM was less than 16 m, which was sufficient for our intended purpose (CGIAR-CSI 2019). In addition to the elevation, we derived other terrain variables (aspect, hill-shade and slope) using the "terrain analysis" plugin in QGIS (QGIS Development Team 2019). The influence of elevation, slope, hill-shade and aspect on soil type, soil moisture content, soil fertility, soil temperature and runoff among other factors were anticipated to influence the occurrence and propagation of Striga weed. Striga has been reported by other studies to be tolerant of a wide range of altitudes from sea level to ~2 480 m a.s.l (Cochrane and Press 1997).

Vegetation phenological variables. Vegetation phenological variables were estimated from 250 m, MODIS 16-day enhanced vegetation index (EVI) time-series composites. We used a 6-year observation period between 2012 and 2018. We computed the vegetation phenological variables using the TIMESAT software (Eklundh and Jönsson 2017; Jönsson and Eklundh, 2002, 2004). TIMESAT enumerates phenological harmonics that occur within a time-series satellite dataset by superimposing localized equations to the time-series data points. Curve smoothing functions are thereafter applied to the model to extract the seasonal vegetation phenological parameters from these vast multi-temporal data dimensions. This consecutively reduces the influence of residual signal noise produced by the variation in the raw EVI time-series data (Hentze, Thonfeld, and Menz 2016; Makori et al. 2017). For this study, we employed a Savitzky-Golay filter to smoothen the fitting curves and removed outliers using a 3- and 5-point window over 2 fitting steps, 3.0 adaptation strength without spike or amplitude cutoffs, a 0.0 season cutoff and a 20% threshold for the beginning and end of the season following a procedure described in Makori et al. (2017). Using this protocol, we extracted 11 vegetation phenological variables (Table 3). For a detailed explanation of the calculated variables and how the TIMESAT algorithm operates, readers are referred to Eklundh and Jönsson (2017) for elaborate information on TIMESAT variables.

Collinearity test of variables used in the Ecological Niche Modeling (ENM)

We used 2-stage variable elimination criteria using the variance inflation factor (VIF) and the Person correlation coefficient. VIF detects multicollinearity by taking each predictor and regressing it against the other variables in a multiple linear regression analysis (Plant 2012). The resulting R^2 value obtained from this multiple regression analysis is then replaced in the VIF calculation formula as shown in Equation 1.

$$VIF_i = \frac{1}{1 - R_i^2} \tag{1}$$

Where *i* is the predictor

In this study, we chose to use the "vifcor" function in the "usdm" package available in R (Naimi et al. 2014; R Core Team 2019). The "vifcor" function iteratively selects pairs of variables with high linear correlation, then eliminates the one with the highest VIF. We set the threshold at *th* = 0.7, which represents a Pearson correlation coefficient (r ≥ 0.7) following Kyalo et al. (2018) recommendation. In principle, a VIF value greater than 10 is evidence of the collinearity problem within a model (Dormann et al. 2013). Although some of the variables from the VIF calculation process showed low VIF values, the correlation matrix (Figure 3) revealed further correlations among some of the variables. Therefore, from the variables with low VIF, we selected the variables that have been reported in the literature to

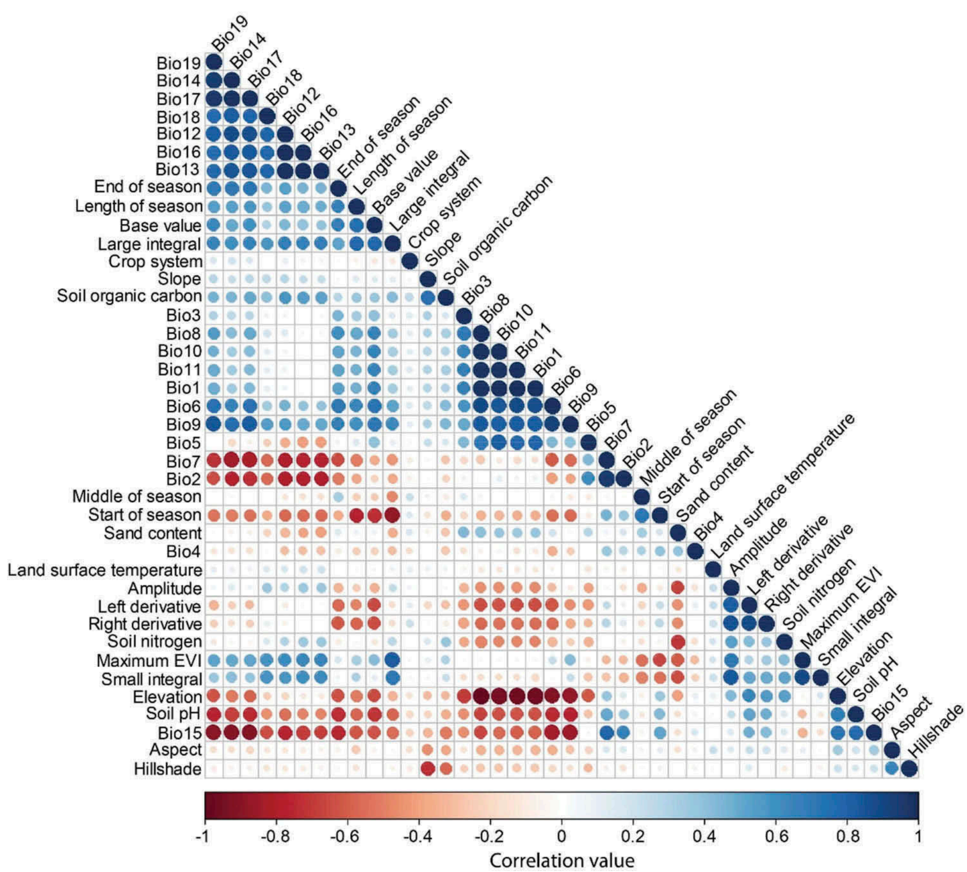


Figure 3. Collinearity matrix for ecological niche models' predictor variables. Darker shades of blue and red color indicate high variable collinearity, while lighter shades indicate low collinearity. Similarly, the smaller the circle, the lower the correlation value.

be of ecological significance (Mandumbu 2017). Our variable elimination procedure resulted in a selection of 21 eligible variables from 40 bioclimatic and remotely sensed variables. The 21 variables that were used in the final modeling procedure are highlighted in bold in Tables 2 and 3.

Collinearity amongst the predictor variables in most ENM causes instability and volatility of the model parameterization and performance (Dormann et al. 2013). The variables correlation matrix using the Pearson correlation coefficient is shown in Figure 3.

Species distribution models implementation

We built Striga occurrence predictive models using the "sdm" package (Naimi and Araújo 2016) performed in R (R Core Team 2019). We used the 50 "presence-only" points data that we collected in the field against 1 000 pseudo-absence points generated using the "sdmdata" function inherent in the "sdm" package. Stockwell and Peterson (2002) concluded that when using ML methods for species niche predictions, the accuracy for predicting the occurrence of a species at a location, was 90% of maximum within 10 sample points, and was near maximal at 50 data points. Therefore, our Striga "presence-only" sample size (i.e. 50) was within the sufficient sample size required for accurate predictions when an ENM is employed at a national scale (Stockwell and Peterson 2002). Often, obtaining real "presence-absence" data is logistically impractical; however like other ENMs, "sdm" allows for the use of background pseudo-absence data (Guan et al. 2020). The package "sdm" combines diverse executions of ENMs (n = 15) within a single platform and uses the same "presence-only" and pseudo-absence data by applying an object-oriented reproducible and extensible framework for ENM in R (Naimi and Araújo 2016). In the present study, we selected and inter-compared only 6 of the 15 modeling techniques in "sdm" as follows: CART, BRT, RF, FDA, GLM, and SVM.

The CART model grows a single decision tree based on the binary partitioning algorithm, which splits the data until it is homogenous, using a hierarchical structure and regression tree (Breiman et al. 1984). Similarly, the BRT model uses the same decision tree approach but improves from the use of a single regression tree by combining multiple decision trees in a process called boosting (Elith, Leathwick, and Hastie 2008). On the other hand, the RF uses these multiple decision trees and randomly grows a forest of decision trees, then predictions are conducted through majority voting for the class with the highest number of votes among these multiple grown trees (Bangira et al. 2019). FDA is a non-parametric multiple regression and additive technique and the GLM uses a linear regression approach (Nelder and Wedderburn 1972), while SVM uses a hyperplane to estimate the divergence of class groupings for the prediction (Hastie, Tibshirani, and Buja 1994; Vapnik 1979). These six algorithms were selected in this study because they are widely used in conducting complex output predictions with relatively high modeling accuracies for regression and classification (Abdel-Rahman, Ahmed, and Ismail 2013; Abdel-Rahman et al. 2016; Makaya et al. 2019; Mosomtai et al. 2016; Tesfamichael et al. 2018). A summary of these models' execution syntax and their corresponding packages used by "sdm" in the parallel model simulations is provided in Table 4.

An ensemble projection approach was used to harmonize the variations produced by the different model predictions. Ensembles have been reported to have superior predictive performance as compared to individual models (Hao et al. 2019). The ensemble modeling fits and maximizes the prediction accuracy with higher reliability as it binds together the highest performance of all the models that have the most acceptable precision and accuracy. In the present study, the function "ensemble" within the "sdm" package was used to harmonize the results of our Striga occurrence prediction amongst our six modeling algorithms using the true skill statistic (TSS) weighted average approach (Naimi and Araújo 2016). Compared to using the most intuitive approach, which applies the mean or median, the weighted average improves the model's predictive ability (Naimi and Araújo 2016; Jafarian, Kargar, and Bahreini 2019). However, the weighted average requires validation of the selected modeling algorithms before inclusion in the "sdm" (Hao et al. 2019). For the present study, we set the threshold to TSS = 0.7 for the models to

Table 4. R software packages used by "sdm" in the parallel execution of the six models; namely (a) boosted regression trees (BRT), (b) classification and regression trees (CART), (c) flexible discriminant analysis (FDA), (d) generalized linear model (GLM), (e) random forest (RF) and (f) support vector machines (SVM).

Algorithm	Syntax code in "sdm"	Package used	Reference
Boosted regression trees	"brt"	gbm	Elith, Leathwick, and Hastie (2008)
Classification and regression trees	"cart"	rpart	Breiman et al. (1984)
Flexible discriminant analysis	"fda"	earth	Hastie, Tibshirani, and Buja (1994)
Generalized linear regression	"glm"	glmnet	(Friedman, Hastie, and Tibshirani (2010)
Random forest	"rf"	randomForest	Liaw and Weiner (2002)
Support vector machines	"svm"	Kernlab	Karatzoglou et al. (2004)

qualify for inclusion in the ensemble as generally a TSS score of > 0.7 points to high agreements between the predictor variable and the independent data (Allouche, Tsoar, and Kadmon 2006).

For consistency, we used the same approach to perform the Striga occurrence predictions for both the current and future climate scenarios. Since our study was more focused on the influence of climate change on the distribution of Striga, we only varied the selected climatic variables, but all the other variables were assumed to remain constant in the future and if they would be differences, we assumed that they would not to be significant enough to cause major variances to the results we obtained.

We further calculated the predicted suitable area for the probability of Striga occurrence for both the current and future scenarios using an average thresholding value of all the models. Using this value, we created a binary raster image of the presence (occurrence) and absence classes for the whole study area. We used the total number of pixels to estimate the total coverage of the predicted area against the unsuitable area.

Models' accuracy validation

The accuracy and variable importance of the models were tested using a 10-fold cross-validation approach. We used the relative variable contribution to the model using the inbuilt randomly split "independent test data set" option available in the "sdm" package. This was automated to universally apply to each of the six models. We further measured the performance of the six models using the receiver operating curve (ROC) by analyzing the area under the curve (AUC) and TSS(Allouche, Tsoar, and Kadmon 2006; Guan et al. 2020). The ROC is a graphical representation of how well the model fits the data points. The values for the AUC range between 0 and 1. Models whose predictions are 100% inaccurate have an AUC of 0, while those with perfect prediction have an AUC of 1. In general, AUC values ≥ 0.7 demonstrate high model prediction performances (Mohammadi et al. 2019). On the other hand, TSS (Equations (2)–(4)) combines sensitivity and specificity to explain the model commission and omission errors (Kyalo et al. 2018). The values of TSS range between −1 to +1, where +1 demonstrates a perfect agreement between the observed and the predicted Striga occurrence, while values ≤ 0 indicates no agreements or that most of the predictions for the Striga occurrence were produced by chance (Allouche, Tsoar, and Kadmon 2006). We, therefore, used the weighted average of the TSS to perform the ensemble predictions. We chose TSS since it is a relatively reliable measure instead of the AUC and chi-squared (X^2) statistics which are somewhat biased and highly sensitive to the proportional extent of the predicted presence observations (Kyalo et al. 2018). The ensemble model merges the strengths of these ENM approaches while minimizing their weaknesses (Araújo et al. 2019; Guan et al. 2020)

$$\mathrm{TSS} = Sensitivity + Specificity - 1 \tag{2}$$

$$Sensitivity = \frac{a}{a+b} \tag{3}$$

$$Specificity = \frac{d}{c+d} \tag{4}$$

where a is true positive, b is a false negative, c is false positive, and d is true negative.

The output maps from the six models and their respective ensembles were imported into a geographical information system (GIS) environment for further analysis. Based on a suggestion by Abdelaal et al. (2019), we reclassified our probability maps into five classes of Striga probability of occurrence. These classes were: (i) very low probability (≤ 0.05), (ii) low probability (0.051–-0.10), (iii) moderate probability (0.11–0.30), (iv) high probability (0.31–0.50), and (v) very high probability (≥ 0.50).

Results

Models' accuracy, comparison, and validation

The VIF statistic of the predictor variables that were included in the modeling approach using the six models is summarized in Tables 2 and 3. The lowest values of VIF were related to remotely sensed variables i.e. end of season (1.02), start of season (1.04), LST (1.07) and amplitude (1.09), while bioclimatic variables had higher values of VIF such as Bio5 (8.21), Bio1 (7.30), Bio18 (5.41) and Bio15 (4.32). However, these values were not large enough to warrant these variables to be eliminated from the modeling. On the other hand, the VIF values for Bio6 and Bio7 were 10.23 and 12.63, respectively. These variables had VIF values greater than 10, so they were excluded from our modeling analysis.

Using the ROC, the patterns of the smoothened graphs of the ten replicated ROCs showed that RF and GLM were relatively consistent in their prediction amongst the model replicates compared to the other models (Figure 4). Some of the replicated graphs using different sets of data folds for the CART, FDA and BRT models were below or closer to the one-to-one line (the black dotted line in Figure 4).

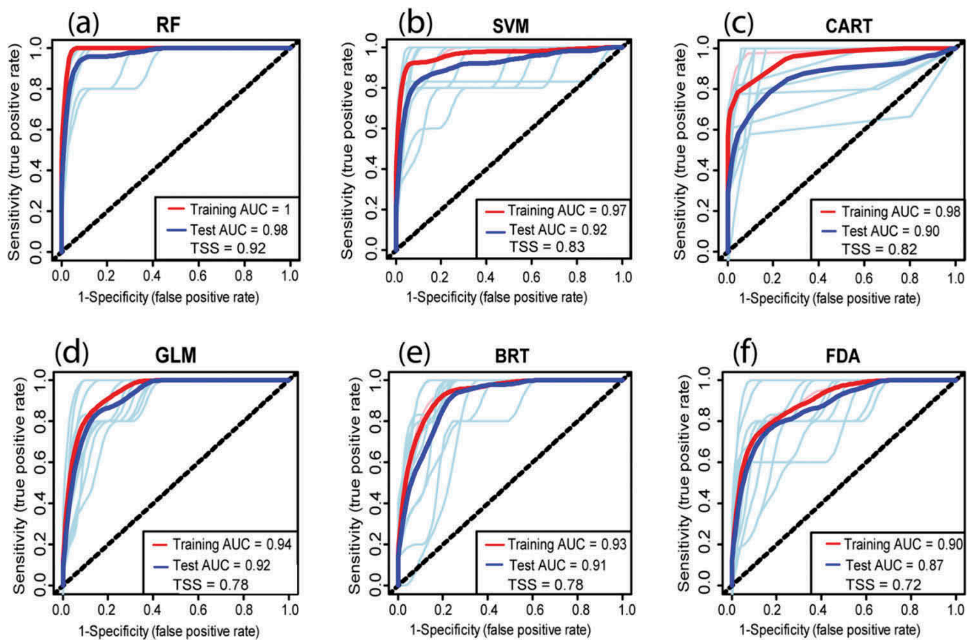


Figure 4. Results of the receiver operating curve (ROC) for the six machine learning and ecological niche model algorithms used to predict Striga occurrence in Zimbabwe; namely: (a) random forest (RF), (b) support vector machines (SVM), (c) classification and regression trees (CART), (d) generalized linear model (GLM), (e) boosted regression trees (BRT), and (f) flexible discriminant analysis (FDA). The red curve represents the smoothened mean area under the curve (AUC) using the training data, while the blue curve depicts the smoothened mean AUC using the test data from the 10-fold cross-validation sampling. The cyan curves show the 10-fold replicated model runs using the training data.

All the models generally showed relatively high accuracy in predicting Striga occurrence in Zimbabwe, with all the models producing acceptable accuracies *viz*., AUC > 0.85 and TSS > 0.70. Further, the models' predictive performance, as indicated by AUC and TSS values revealed that RF had the highest values of AUC (0.98) and TSS (0.92) (Figure 4). The FDA model produced the lowest AUC (0.87) and TSS (0.72) scores. We, however, observed variations where models such as the GLM had a higher AUC, but lower TSS in comparison to other models such as CART. Nonetheless, the FDA performed the least using the AUC and the TSS accuracy measures.

Variable importance using the current climate scenario

A total of 5 out of the 21 predictor variables appeared in the ten most relevant variables for all six models. The five predictor variables regarded as very relevant by all the models are "base value," "start of season," "temperature seasonality" (Bio4), "maximum temperature of the warmest month" (Bio5), and "precipitation seasonality" (Bio15). The respective variable contributions in the different models are summarized in Figure 5. The Bio5 variable appeared twice as the most relevant variable i.e. for the RF and BRT models, while the Bio1 variable also appeared twice as the most relevant variable for the GLM and FDA models. Bio4 and Bio15 were also selected as important predictors for the CART and SVM models, respectively.

Further analysis of the variable importance revealed that the bioclimatic and seasonality parameters dominated the most relevant list while the edaphic, terrain, LST and cropping system were not particularly relevant across the six algorithms tested. Terrain variables appeared in all the models at different contribution levels; however, elevation appeared more frequently than the other terrain variables. Similarly, regarding the edaphic variables, soil organic carbon and sand content dominated their category with varying contributions across the six models. The cropping system variable appeared once under the BRT model, however, with a very low contribution to the entire model. LST did not appear among the ten most important variables for the six models. RF, which had the highest accuracy (AUC = 0.98) amongst the other models, selected Bio5, Bio4, and Bio15 as the most relevant predictor variables for the estimating occurrence probability of Striga in Zimbabwe (Figure 5).

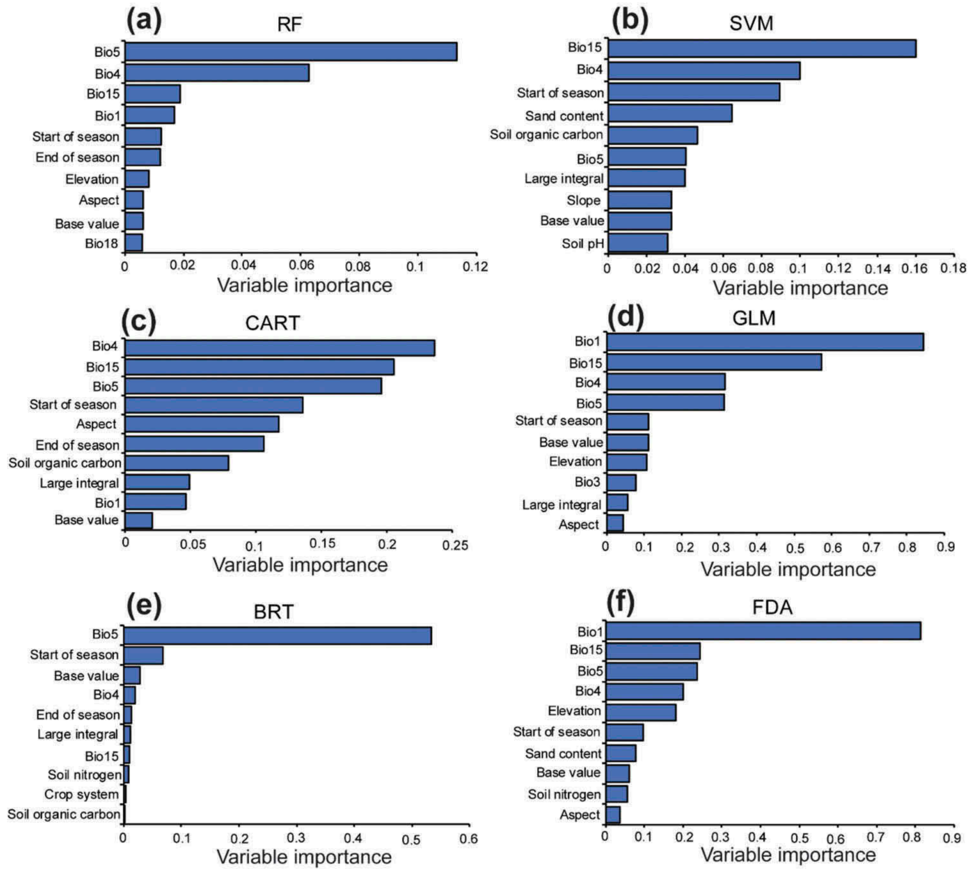


Figure 5. The ten most important variables for the six ecological niche model algorithms used to predict Striga occurrence in Zimbabwe; namely (a) random forest (RF), (b) support vector machines (SVM), (c) classification and regression trees (CART), (d) generalized linear model (GLM), (e) boosted regression trees (BRT), and (f) flexible discriminant analysis (FDA).

Ecological niche models for predicting the occurrence of Striga using the current climate scenario

The six ENMs using the 21 predictor variables exhibited varied results for predicting Striga probability of occurrence (Figure 6). However, all six models predicted the ecological niche and Striga occurrence to be within the central plateau (mainly ecological region II, III, and IV) of the country's main watershed as shown by the warmer colors (yellow, orange and red) in Figure 6. Areas close to the boundaries of Zimbabwe (ecological region V) represented by the cooler colors (green) were predicted to be relatively safe and free from potential Striga infestation using SVM and CART, while in the eastern highlands of Zimbabwe (i.e. ecological region I and II), the occurrence of Striga was predicted using RF, FDA, and BRT.

Ensemble projection for predicting the occurrence of Striga using the current climate scenario

The results of the ensemble projection of the six models using the current climate scenario (1950–2000) combined the best predictions of all the models and estimated the overall Striga probability of occurrence (Figure 7). The highest prevalence and probability of occurrence was predicted to be in the Midlands and Masvingo provinces which are in regions III and IV that are regions with very low intensity of irrigation agriculture. However, some relatively similar predictions were also observed in Matabeleland North province toward the Kariba dam which has climate characteristics of ecological region III. Similarly, parts of the provinces of Manicaland (ecological region I and II), Bulawayo (ecological region IV) and Mashonaland East (ecological region II and III) exhibited moderate, high to very high probabilities of potential Striga incidences. The highest probability of occurrence was observed in agro-ecological regions I, II, III, and IV, whereas very little to none Striga probabilities of occurrence were predicted in region V. Interestingly, the ensemble model was precise (AUC = 0.98) in predicting Striga occurrence following the boundaries of region V where the Striga occurrence is predicted to be very low. The area toward the west of region IV was predicted to have very low Striga probability of occurrence, whereas the central and eastern areas within region IV were predicted to have high to very high incidences of Striga

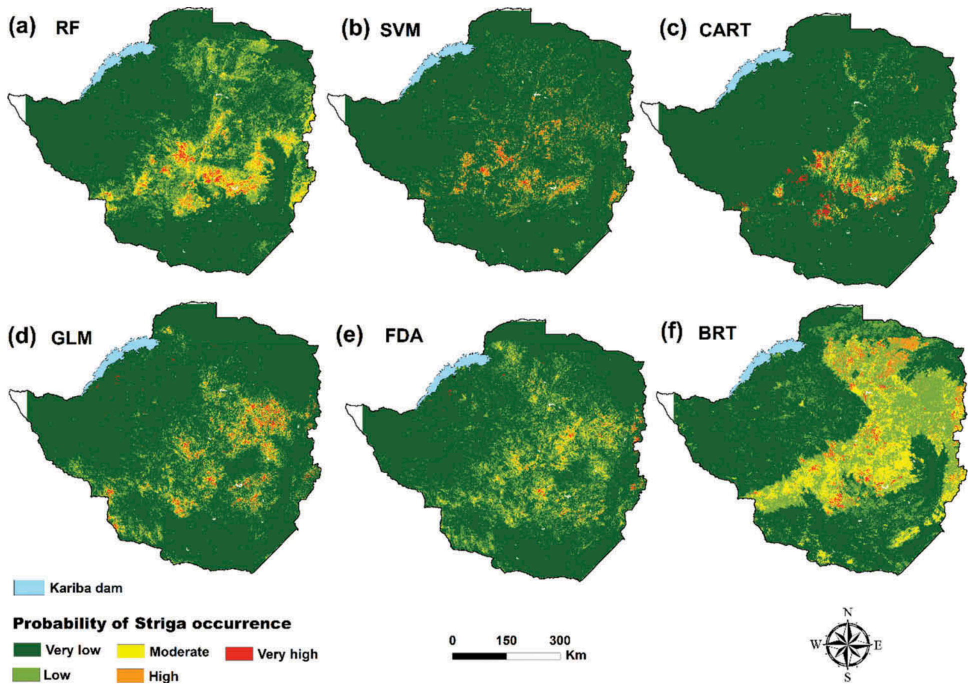


Figure 6. Striga probability of occurrence using the current remotely sensed and bioclimatic variables and the six ecological niche model algorithms: (a) random forest (RF), (b) support vector machines (SVM), (c) classification and regression trees (CART), (d) generalized linear model (GLM), (e) flexible discriminant analysis (FDA), and (f) boosted regression trees (BRT).

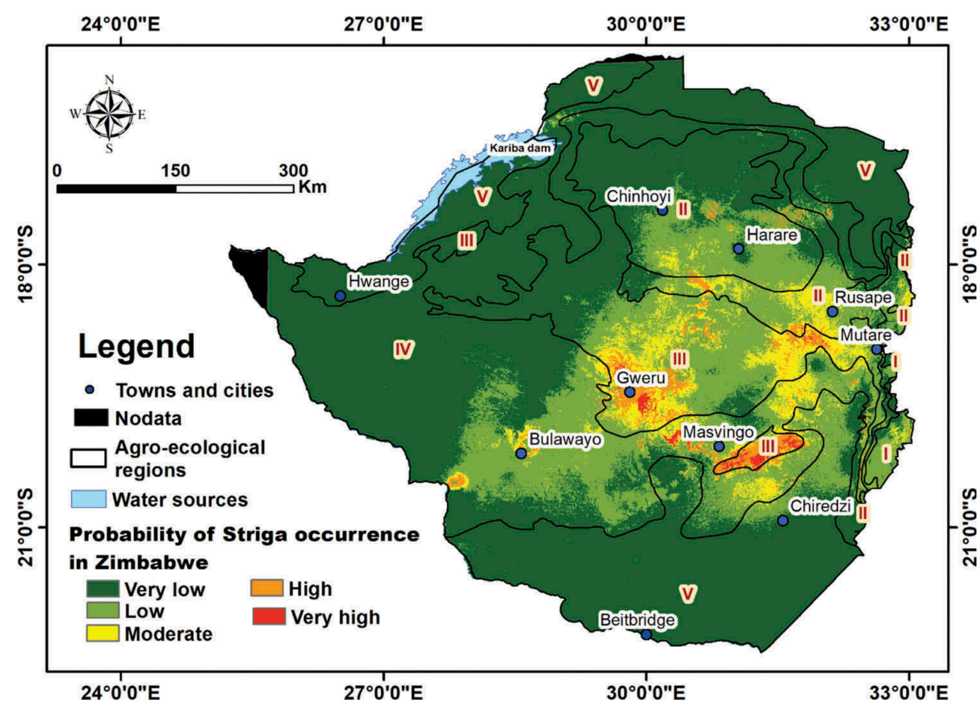


Figure 7. Current Striga probability of occurrence predicted using ensemble projection and the weighted average of the true skill statistic (TSS) of the six prediction models, *viz*. random forest, support vector machines, classification and regression trees, generalized linear model, flexible discriminant analysis, and boosted regression trees ecological niche model algorithms.

occurrence. In general, the warmer colors also showed that the Striga probability of occurrence is skewed toward the central and eastern regions of the country with relatively high-altitudes (800 m–1 600 m a.s.l) compared to the low-altitude areas (< 800 m a.s.l) on the west, south, and north (Figure 7).

Comparison of the ensemble predictions using the current and future climate scenarios

We detected slight differences in the suitable area between the current and future climate scenarios for Striga occurrence in Zimbabwe. We observed that Striga occurrence would shift toward the North, i.e. Mashonaland West and East, which are in ecological region II (Figure 8) and will be reduced on the southern regions of the country, i.e. Matabeleland North and South occurring in ecological region IV (Figure 8). The future model predicted a very high increase in the area that shall be suitable for Striga, particularly for Masvingo and Midlands provinces which are in the ecological region III. We noted that the intensity of the severity as evidenced by the increase of most areas from the moderate class to very high probability was particularly in ecological region III.

The current area suitable for Striga occurrence in Zimbabwe is 7.4% of the total area, while an increase of ~ 0.73% is likely by the end of 2050 due to climate change. Therefore, the approximate area currently suitable for the occurrence of Striga is 28 916 km^2, while we expect it to increase to 31 768 km^2 (8.13%) by the year 2050 using the maximum carbon emission scenario (RCP8.5). Our estimated increase of the area occupied by Striga due to climate change by the year 2050 is 2 852 km^2 which is an estimated gradual increase rate of ~95 km^2/yr^{-1} over 30 years.

Discussion

In this study, we used six ML and ENM models to predict the current Striga probability of occurrence in Zimbabwe. We followed the best practice standards for ENM by assessing the quality of the response variables, predictor variables, model evaluation ideals and building multiple models using the same data following the protocols suggested by Araújo et al. (2019). Although we used a sufficient sample size required for accurate predictions when using ENMs at a national scale (Stockwell and Peterson 2002), the performance of some models like FDA and GLM which require a relatively large sample size could have been reduced.

Model performances

Generally, predictive models with AUC and TSS values larger than 0.7 suggest plausible predictive and simulation performance (Elith et al. 2010). In our study, AUC and TSS values for all the six models as shown by the ROCs, were above the 0.7 threshold, demonstrating that the models performed well in simulating the distribution of Striga in Zimbabwe. As expected, the model accuracies and the predicted areas differed across the six models, since models depend on different mathematical functions and tuning parameters (Araújo et al. 2019). Using AUC and TSS, our results pointed to RF as the best predictive model, which was consistent with our hypothesis. Based on the obtained AUC and TSS results from this study, we recommend the use of the ensemble, RF, SVM and CART for Striga predictive modeling using multi-source data. These recommended methods have also been used and suggested by many researchers as the best for simulating predictions for invasive weed species occurrence and mapping their geographical

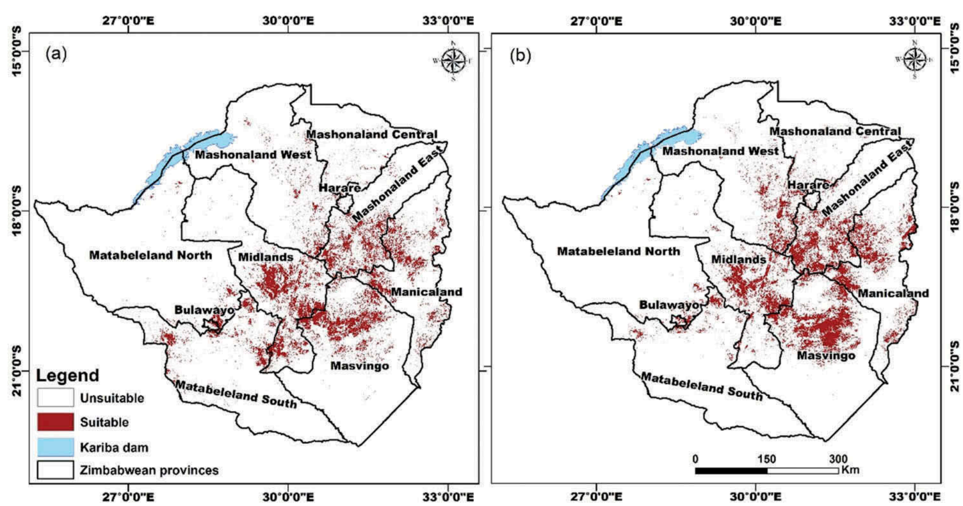


Figure 8. Striga probability of occurrence predicted using the representative concentration pathway (RCP:8.5), ensemble projection and the weighted average of the true skill statistic (TSS) of six ecological niche model algorithms, *viz*., random forest, support vector machines, classification and regression trees, generalized linear model, flexible discriminant analysis, and boosted regression trees. (a) Current (1950–2000) and (b) future (2041–2060) climate scenario.

niches (Mudereri et al. 2019a; Tesfamichael et al. 2018; Landmann et al. 2020; Guan et al. 2020).

Importantly, we, however, noted that there were huge overlaps and similarities in the areas anticipated to be suitable for Striga occurrence. These varied outputs and accuracy results are in agreement with other studies that have used multiple models approach in ENM (Hao et al. 2019; Jafarian, Kargar, and Bahreini 2019; Mohammadi et al. 2019; Guan et al. 2020). Jafarian, Kargar, and Bahreini (2019) used four predictive models to simulate the occurrence of five dominant plant species in Iran and concluded that the ensemble method yielded high predictive power compared to the individual models. On the other hand, Mohammadi et al. (2019) also compared MaxEnt and "sdm" to predict two rodent species and they established that all models were comparable and demonstrated high predictive power. Similarly, in our study, there is no convincing indication to prove that one model is significantly better than the other. Therefore, regarding the future investigations that will focus on the accuracy of Striga occurrence and prediction, it is recommended to include several models in an ensemble approach to reduce the modeling uncertainties.

Striga probability of occurrence in the current climate scenario and under climate change

Fundamentally, input data preparation is key to determine and improve the accuracy and dependability of the outputs derived from predictive models (Araújo et al. 2019). ENMs reflect the deep interrelationships and interactions among species and their environmental parameters. Using the package "usdm" and the "vifcor" function provided an easy and practical way of eliminating the correlated variables systematically (Jafarian, Kargar, and Bahreini 2019). Specifically, the use of VIF as a measure of collinearity and elimination of conflating variables improved the accuracy of our modeling results. This is per other studies that have successfully employed VIF to select a few noncorrelated predicted variables (Muposhi et al. 2016; Abdelaal et al. 2019). The non-conflating variables (n = 21) that were finally used in the modeling experiments were crucial in explaining the occurrence of the Striga weed. Notwithstanding, the important variables that were selected by our models for mapping Striga occurrence were local and not global. That means the variables were only relevant for modeling Striga in Zimbabwe and not somewhere else on the globe. As anticipated, our results showed that the interrelationship between temperature (Bio1, Bio4, and Bio5) and precipitation (Bio15) was central in defining the ecological niche of the Striga weed. This concurred with the results reported by Cotter and Sauerborn (2012), who alluded to the variation in the current and future distribution of Striga in the entire African continent to be influenced by mean annual temperature (Bio1). However, our future Striga prediction models should be interpreted with some caution as they were not yet validated.

Striga requires both optimum rainfall and temperature for germination, growth, propagation and simultaneously the growth of its hosts i.e. cereal crops (Mandumbu et al. 2017a). However, extreme temperature and heavy rainfall conditions limit the propagation of Striga (Rich and Ejeta 2007); hence, the very low probability of Striga occurrence in ecological region V of Zimbabwe. Region V, which is mostly on the borders of Zimbabwe, experiences very high temperatures and low rainfall making it unsuitable for crop production, hence the unavailability of cereals that are Striga hosts. However, with the increase in temperatures anticipated through climate change, most farmers in all agro-ecological regions of Zimbabwe are likely to shift to planting C_4 crops like sorghum and millet which are drought-resistant but are attractive to the occurrence of Striga (Mandumbu 2017). Agro-ecological regions I–IV experience very high to moderate rainfall and temperature compared to the ecological region V (Mugandani et al. 2012). This could have been the reason for the high prediction of Striga occurrence in these regions. These regions have varied climatic conditions, but our modeling and mapping results showed that ecological regions II, III, and IV have the most optimum climatic conditions for the germination, growth, reproduction, and spread of Striga species. Because of the immense dependence on the distribution of Striga on climatic variables, future climate conditions may greatly determine the suitable niche for Striga (Mandumbu et al. 2017a; Cotter and Sauerborn 2012).

Because of climate change, the increase in carbon dioxide and temperature changes are likely to lead to an increase in the germination and spread rates of Striga in areas that were once non-Striga occurrence areas. Our results agreed with the perception reported by Mandumbu et al. (2017a), who argued that the future increase in temperature would increase the breaking of dormancy of Striga seeds, thereby increasing its germination rate. As could be seen from our results, Striga shall occupy new adjacent areas to the already infested areas, mostly in regions III and IV of Zimbabwe. These areas are predicted to have temperature ranges between 20° C and 35° C. This phenomenon is likely to result in increased areas occupied by Striga and enhance the intensity and severity of the crop losses caused by this weed. Additionally, as the temperature continues to increase in the future, crops that are currently not affected by Striga such as the winter wheat may eventually become susceptible to the weed infestation

(Mandumbu et al. 2017a). Therefore, any efforts targeted at curbing the spread of Striga in the future should focus on areas with the likelihood of temperature increase and a reduced amount of rainfall.

Soil N is reported to constrain the germination of Striga seeds by reducing the production of Strigolactones, the chemical that is produced by the host plants to simulate the germination of Striga seeds, while soil N also increases the vegetative growth of the host plant (Ekeleme et al. 2014). Notwithstanding, our results indicated that the edaphic factors that we tested (i.e. soil N, soil pH, soil organic C, and soil sand content) had little effects on our Striga modeling accuracy, probably because of the interplay between them and precipitation (i.e. Bio15). To the best of our knowledge, soil moisture affects nutrient motility, particularly nitrogen, which is mostly unstable and susceptible to leaching (Yoneyama et al. 2007). It is worth noting that precipitation could be a proxy for soil moisture that is an important predictor variable for triggering the germination of Striga seeds. It is expected that the edaphic factor variables and precipitation are intimately interlinked. Hence, they tend to result in low performance of each other in predictive modeling, due to the relatively high correlation. However, various studies argue that degraded soils of high acidity promote the growth and proliferation of Striga weed (Midega et al. 2017; Larsson 2012; Yoneyama et al. 2007). On the other hand, it is reported that the degradation of soils and increasing its soil pH is expected to worsen with the change in climate (Mandumbu et al. 2017b). As earlier mentioned, in response to the low soil fertility and drought, farmers are likely to shift to C_4 plants which are more tolerant of droughts and high temperatures but are more susceptible to the Striga infestation.

Our study also shows that "start of season" and EVI "base value" were among the most important predictor variables in all the six ENM algorithms. Striga depends on the availability of the host cereal crops for its germination, survival, and propagation. The minimum level of greenness in the whole season and the level of greenness at the start of the season can be described by the EVI values which foretell crop planting date and crop health (e.g. Striga infestation rate). Similarly, the minimum value of EVI during the season signifies the crop health status. Thus, the start of season and EVI base value can be very relevant variables to predict the occurrence of Striga in semi-arid environments.

Implications of our study

Modeling the potential distribution of weeds such as Striga is useful in agricultural management systems in areas most likely to be susceptible to invasion and colonization. Our study supports national scale preparedness and management strategies for the protection of key crops from diseases and pests, especially in the face of climate change and variability. Furthermore, results from the present study show that using ENM is one of the most reliable and central tools for determining the fundamental and realized niche of Striga within a geographical space. Our study showed that Striga spread and propagation is likely to be within the adjacent areas in the ecological region III of Zimbabwe. Although this cannot precisely be empirically derived from our current models, we infer and anticipate that wind, water, animal and human movement shall be the main modes of Striga seed spread. Striga plants are highly productive with the potential to produce between 10 000 and 200 000 seeds per plant (Ejeta and Butler 1993). These seeds are of lightweight (~ 4–7 µg per seed), which makes them easy to disperse by wind water or animals (Mandumbu et al. 2017a; Wan and Wang 2019). Similarly, farmers within the same or adjacent areas are likely to exchange farming equipment, thereby promoting the spread of Striga seeds. Additionally, since the soils in the ecological region III of Zimbabwe are deep Kalahari sands, we anticipate further degradation of these soils which leads to losses in soil fertility and ultimately could promote the spread of Striga (Yoneyama et al. 2007). To combat Striga occurrence and spread particularly in regions II, III and IV of Zimbabwe, our study can be utilized for guiding the implementation and upscaling of "push-pull" technology (PPT). PPT is a climate-smart integrated farming system that uses the legume *Desmodium* as an intercrop to combat the reproduction cycle of Striga and repel insect pests i.e. stemborers and fall armyworm (Khan et al. 2014). *Desmodium* secretes a set of compounds that promote the suicidal germination of Striga and effectively inhibits the possibilities of the Striga to attach their roots to the roots of the host plant. Interested readers are referred to Khan et al. (2006) and Pickett et al. (2014), for elaborate information on the PPT farming system.

Conclusions

We compared ML ENMs i.e. RF, CART, SVM, BRT, GLM, FDA and their respective ensemble for predicting the probability of Striga occurrence in Zimbabwe using multi-source bioclimatic and remotely sensed data. We established that RF, CART, SVM and the ensemble approach, yield the most accurate Striga occurrence prediction results in Zimbabwe. Our results showed that temperature and precipitation are the key drivers of the occurrence of Striga. In addition, the Striga epidemic in Zimbabwe is highly likely to worsen and spread into new areas where it was not initially found, particularly in the ecological regions I to IV of the country. Therefore, immediate and palliative action is critical to contain and manage the spread and intensity of Striga in Zimbabwe. Our results could help researchers, policymakers, extension officers and various other stakeholders to employ and implement effective and early Striga management options to contain and eradicate the weed. Because our study employed a synoptic approach at a national scale using datasets at a coarse spatial resolution (250 m x 250 m pixel size), future studies should focus on developing localized early warning advisory platforms and high resolution (i.e. submeter) remotely sensed observations to detect and monitor Striga infestation and density. Specifically, the use of unmanned aerial vehicles (UAVs) should be investigated for appropriate use to early detect Striga occurrence and suitable habitats before its flowering stages. Although we used Zimbabwe as a case study, our modeling results can be upscaled in other African countries that possess similar agro-ecological characteristics.

Highlights

- We developed ecological suitability models for *Striga asiatica*, an invasive cereal crop weed
- Key input predictors were climatic, edaphic, phenology, terrain and cropping system
- Striga occurrence is more pronounced in the agro-ecological regions I–IV in Zimbabwe
- Striga will spread into formerly non-striga areas because of climate change

Acknowledgements

The authors express their gratitude to the farmers in Masvingo and Midlands provinces of Zimbabwe for their contribution through information sharing and cooperation. We also extend our gratitude to the Masvingo and Midlands province extension officers under the Ministry of Agriculture in Zimbabwe and members of Kushereketa Rural Development Organization (KURDO) for their support during data collection.

Disclosure statement

The authors declare that there are no conflicts of interest.

Funding

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 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. Bester Tawona Mudereri was supported by a German Academic Exchange Service (DAAD) In-Region Postgraduate Scholarship;Biovision Foundation for Ecological Development (Switzerland) [BV DPP-010/2019]

ORCID

Bester Tawona Mudereri http://orcid.org/0000-0001-9407-7890
Elfatih Mohamed Abdel-Rahman http://orcid.org/0000-0002-5694-0291
Timothy Dube http://orcid.org/0000-0003-3456-8991
Zeyaur Khan http://orcid.org/0000-0002-3548-7563

Data availability

Data are available on request.

References

Abdelaal, M. M., G. Fois, Fenu, and G. Bacchetta. 2019. "Using MaxEnt Modeling to Predict the Potential Distribution of the Endemic Plant Rosa Arabica Crép. In Egypt." *Ecological Informatics* 50: 68–75. doi:10.1016/j.ecoinf.2019.01.003.

Abdel-Rahman, E. M., F. B. Ahmed, and R. Ismail. 2013. "Random Forest Regression and Spectral Band Selection for Estimating Sugarcane Leaf Nitrogen Concentration Using EO-1 Hyperion Hyperspectral Data." *International Journal of Remote Sensing* 34 (2): 712–728. doi:10.1080/01431161.2012.713142.

Abdel-Rahman, E. M., T. Landmann, R. Kyalo, G. Ong'amo, S. Mwalusepo, S. Sulieman, and B. Le Ru. 2016. "Predicting Stem Borer Density in Maize Using RapidEye Data and Generalized Linear Models." *International Journal of Applied Earth Observation and Geoinformation* 57. doi:10.1016/j.jag.2016.12.008.

Allouche, O., A. Tsoar, and R. Kadmon. 2006. "Assessing the Accuracy of Species Distribution Models: Prevalence, Kappa and the True Skill Statistic (TSS)." *Journal of Applied Ecology* 43 (6): 1223–1232. doi:10.1111/j.1365-2664.2006.01214.x.

Araújo, M. B., R. P. Anderson, M. A. Barbosa, C. M. Beale, C. F. Dormann, E. Regan, R. A. Garcia, et al. 2019. "Standards for Distribution Models in Biodiversity Assessments." *Science Advances* 5 (1): 1–12. doi:10.1126/sciadv.aat4858.

Ayebare, S., R. H. Mdegela, L. Nyakarahuka, S. A. Ndimuligo, E. R. Mwakapeje, H. E. Nonga, E. Skjerve, and G. Mosomtai. 2018. "Ecological Niche Modeling as a Tool for Prediction of the Potential Geographic Distribution of Bacillus Anthracis Spores in Tanzania." *International Journal of Infectious Diseases* 79. International Society for Infectious Diseases: 142–151. doi:10.1016/j.ijid.2018.11.367.

Bangira, T., A. Van Niekerk, M. Menenti, and S. M. Alfieri. 2019. "Comparing Thresholding with Machine Learning Classifiers for Mapping Complex Water." *Remote Sensing* 1–21. doi:10.3390/rs11111351.

Breiman, L., J. Friedman, C. J. Stone, and R. A. Olshen. 1984. "Classification Algorithms and Regression Trees." *Wadsworth*

Statistics/Probability, 246–280. https://rafalab.github.io/pages/649/section-11.pdf

Breiman, L. 2001. “Random Forests.” *Machine Learning* 45: 5–32. https://link.springer.com/content/pdf/10.1023%2FA%3A1010933404324.pdf

CGIAR-CSI. 2019. “SRTM.” http://srtm.csi.cgiar.org/

Cian, F., M. Marconcini, and P. Ceccato. 2018. “Normalized Difference Flood Index for Rapid Flood Mapping: Taking Advantage of EO Big Data.” *Remote Sensing of Environment* 209: 712–730. doi:10.1016/j.rse.2018.03.006.

Cochrane, V., and M. C. Press. 1997. “Geographical Distribution and Aspects of the Ecology of the Hemiparasitic Angiosperm Striga Asiatica (L.) Kuntze : A Herbarium Study.” *Journal of Tropical Ecology* 13 (3): 371–380. doi:10.1017/S0266467400010579.

Cotter, M., and J. Sauerborn. 2012. “Understanding the Present Distribution of the Parasitic Weed Striga Hermonthica and Predicting Its Potential Future Geographic Distribution in the Light of Climate Change.” *Julius -Kühn-Archiv* 13 (15): 630–636. doi:10.5073/jka.2012.434.082.

Csillag, F., Á. Kummert, and M. Kertész. 1992. “Resolution, Accuracy and Attributes: Approaches for Environmental Geographical Information Systems.” *Computers, Environment and Urban Systems* 16 (4): 289–297. doi:10.1016/0198-9715(92)90010-O.

Degbelo, A., and W. Kuhn. 2018. “Spatial and Temporal Resolution of Geographic Information: An Observation-Based Theory.” *Open Geospatial Data, Software and Standards* 3 (1). Open Geospatial Data, Software and Standards. doi:10.1186/s40965-018-0053-8.

Dormann, C. F., J. Elith, S. Bacher, C. Buchmann, G. Carl, G. Carré, J. R. García Marquéz, et al. 2013. “Collinearity: A Review of Methods to Deal with It and A Simulation Study Evaluating Their Performance.” *Ecography* 36 (1): 027–046. doi:10.1111/j.1600-0587.2012.07348.x.

Ejeta, G., and J. Gressel. 2007. *Integrating New Technologies for Striga Control: Towards Ending the Witch-Hunt. Integrating New Technologies for Striga Control: Towards Ending the Witch-Hunt*. Singapore: World Scientific. doi:10.1142/6470.

Ejeta, G., and L. G. Butler. 1993. “Host-Parasite Interactions Throughout the *Striga* Life Cycle, and Their Contributions to Striga Resistance.” *African Crop Science Journal* 1(2): 75–80. doi:10.4314/acsj.v1i2.69889

Ekeleme, F., J. M. Jibrin, A. Y. Kamara, M. Oluoch, A. M. Samndi, and A. A. Fagge. 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 Protection* 66. Elsevier Ltd: 90–97. doi:10.1016/j.cropro.2014.09.001.

Eklundh, L., and P. Jönsson. 2017. *TIMESAT 3.3 With Seasonal Trend Decomposition and Parallel Processing Software Manual*, 1–92. Sweden: Lund and Malmo University.

Elith, J., J. R. Leathwick, and T. Hastie. 2008. “A Working Guide to Boosted Regression Trees.” *Journal of Animal Ecology* 77 (4): 802–813. doi:10.1111/j.1365-2656.2008.01390.x.

Elith, J., T. Hastie, M. Dudík, Y. En Chee, C. J. Yates, and S. J. Phillips. 2010. “A Statistical Explanation of MaxEnt for Ecologists.” *Diversity and Distributions* 17 (1): 43–57. doi:10.1111/j.1472-4642.2010.00725.x.

FAO, and ACFD. 1999. *A Fertilizer Strategy for Zimbabwe. Food and Agriculture Organization of the United Nations(FAO)*. Rome: FAO.

Fisher, R. A. 1936. “The Use of Multiple Measurements in Taxonomic Problems.” *Annals of Eugenics* 7: 179–188. doi:10.1111/j.1469-1809.1936.tb02137.x.

Friedman, J. 2001. “Greedy Function Approximation: A Gradient Boosting Machine.” *The Annals of Statistics* 29 (5):1189-1232.

Friedman, J., T. Hastie, and R. Tibshirani. 2010. “Regularization Paths for Generalized Linear Models via Coordinate Descent.” *JournalofStatisticalSoftware* 33 (1): 1–22. http://www.jstatsoft.org/v33/i01/10

Guan, B. C., H. J. Guo, S. S. Chen, D. M. Li, X. Liu, X. Gong, and G. Ge. 2020. “Shifting Ranges of Eleven Invasive Alien Plants in China in the Face of Climate Change.” *Ecological Informatics* 55: 101024. doi:10.1016/j.ecoinf.2019.101024.

Guo, Y., X. Li, Z. Zhao, and Z. Nawaz. 2019. “Predicting the Impacts of Climate Change, Soils and Vegetation Types on the Geographic Distribution of Polyporus Umbellatus in China.” *Science of the Total Environment* 648: 1–11. doi:10.1016/j.scitotenv.2018.07.465.

Hao, T., J. Elith, G. Guillera-Arroita, and J. J. Lahoz-Monfort. 2019. “A Review of Evidence about Use and Performance of Species Distribution Modelling Ensembles like BIOMOD.” *Diversity and Distributions* 25 (5): 839–852. doi:10.1111/ddi.12892.

Hastie, T., R. Tibshirani, and A. Buja. 1994. “Flexible Discriminant Analysis by Optimal Scoring.” *Journal of the American Statistical Association* 89 (428): 1255–1270. doi:10.1080/01621459.1994.10476866.

Hengl, T., G. B. M. Heuvelink, B. Kempen, J. G. B. Leenaars, M. G. Walsh, K. D. Shepherd, A. Sila, et al. 2015. “Mapping Soil Properties of Africa at 250 M Resolution: Random Forests Significantly Improve Current Predictions.” *PloS One* 10 (6): 1–26. doi:10.1371/journal.pone.0125814.

Hentze, K., F. Thonfeld, and G. Menz. 2016. “Evaluating Crop Area Mapping from Modis Time-Series as an Assessment Tool for Zimbabwe’s ‘Fast Track Land Reform Programme.’.” *PloS One* 11 (6): 1–22. doi:10.1371/journal.pone.0156630.

IPCC. 2014. *Climate Change 2014. Synthesis Report. Versión Inglés. Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. doi:10.1017/CBO9781107415324.

Jafarian, Z., M. Kargar, and Z. Bahreini. 2019. “Which Spatial Distribution Model Best Predicts the Occurrence of Dominant Species in Semi-Arid Rangeland of Northern Iran?” *Ecological Informatics* 50: 33–42. doi:10.1016/j.ecoinf.2018.12.011.

Jönsson, P., and L. Eklundh. 2002. “Seasonality Extraction by Function Fitting to Time-Series of Satellite Sensor Data.” *IEEE Transactions on Geoscience and Remote Sensing* 40 (8): 1824–1832. doi:10.1109/TGRS.2002.802519.

Jönsson, P., and L. Eklundh. 2004. “TIMESAT - A Program for Analyzing Time-Series of Satellite Sensor Data.” *Computers and Geosciences* 30 (8): 833–845. doi:10.1016/j.cageo.2004.05.006.

Karatzoglou, A., A. Smola, K. Hornik, and A. Zeileis. 2004. “Kenlab - an S4 Package for Kernel Methods in R.” *Journal of Statistical Software* 11 (9): 1–20. doi:10.18637/jss.v011.i09.

Khan, Z. R., C. A. O. Midega, J. O. Pittchar, A. W. Murage, M. A. Birkett, T. J. A. Bruce, and J. A. Pickett. 2014. “Achieving Food Security for One Million Sub-Saharan

African Poor through Push–Pull Innovation by 2020." *Philosophical Transactions of the Royal Society B: Biological Sciences* 369 (1639). The Royal Society: 20120284. doi:10.1098/rstb.2012.0284.

Khan, Z. R., J. A. Pickett, L. J. Wadhams, A. Hassanali, and C. A. O. Midega. 2006. "Combined Control of Striga Hermonthica and Stemborers by Maize-Desmodium Spp. Intercrops." *Crop Protection* 25 (9): 989–995. doi:10.1016/j.cropro.2006.01.008.

Kuri, F., A. Murwira, K. S. Murwira, and M. Masocha. 2018. "Accounting for Phenology in Maize Yield Prediction Using Remotely Sensed Dry Dekads." *Geocarto International* 33 (7): 723–736. doi:10.1080/10106049.2017.1299798.

Kuri, F., M. Masocha, A. Murwira, and K. S. Murwira. 2019. "Differential Impact of Remotely Sensed Dry Dekads on Maize Yield in Zimbabwe." *Geocarto International*. doi:10.1080/10106049.2019.1583774.

Kyalo, R., E. M. Abdel-Rahman, S. Mohamed, S. Ekesi, B. Christian, and T. Landmann. 2018. "Importance of Remotely-Sensed Vegetation Variables for Predicting the Spatial Distribution of African Citrus Triozid (Trioza Erytreae) in Kenya." *ISPRS International Journal of Geo-Information* 7 (11): 429. doi:10.3390/ijgi7110429.

Landmann, T., D. Eidmann, N. Cornish, J. Franke, and S. Siebert. 2019. "Optimizing Harmonics from Landsat Time Series Data: The Case of Mapping Rainfed and Irrigated Agriculture in Zimbabwe." *Remote Sensing Letters* 10 (11): 1038–1046. doi:10.1080/2150704X.2019.1648901.

Landmann, T., Olena D., Gohar G., Jackson K., and Elfatih M. Abdel-Rahman. 2020. "Wide-area Invasive Speciespropagation Mapping Is Possible Using Phenometric Trends." *ISPRS Journal of Photogrammetry and Remote Sensing* 159: 1–12. doi: 10.1016/j.isprsjprs.2019.10.016.

Larsson, M. 2012. "Soil Fertility Status and Striga Hermonthica Infestation Relationship Due to Management Practices in Western Kenya." Swedish University of Agricultural Sciences. http://stud.epsilon.slu.se/4488/1/larsson_m_120704.pdf

Liaw, A., and M. Weiner. 2002. "Classification and Regression by RandomForest." *R News* 2 (3): 18–22. doi:10.1177/154405910408300516.

Makaya, N. P., O. Mutanga, Z. Kiala, T. Dube, and K. E. Seutloali. 2019. "Assessing the Potential of Sentinel-2 MSI Sensor in Detecting and Mapping the Spatial Distribution of Gullies in a Communal Grazing Landscape." *Physics and Chemistry of the Earth* 112: 66–74. doi:10.1016/j.pce.2019.02.001.

Makori, D., O. Mutanga, J. Irungu, G. Mosomtai, J. Odindi, S. Makau, K. Nkoba, et al. 2017. "Predicting Spatial Distribution of Key Honeybee Pests in Kenya Using Remotely Sensed and Bioclimatic Variables: Key Honeybee Pests Distribution Models." *ISPRS International Journal of Geo-Information* 6 (3): 66. doi:10.3390/ijgi6030066.

Mandumbu, R. 2017. "The Stability of Tolerance of Sorghum Spp to Striga Asiatica L. Kuntze under Diverse Conditions and Existence of Pre- Attachment Resistance." *University of Fort Hare South Africa*.

Mandumbu, R., C. Mutengwa, S. Mabasa, and E. Mwenje. 2017a. "Predictions of the Striga Scourge under New Climate in Southern Africa: A Perspective." *Journal of Biological Sciences* 17 (5): 194–201. doi:10.3923/jbs.2017.194.201.

Mandumbu, R., C. Mutengwa, S. Mabasa, E. Mwenje, J. Gotosa, and V. T. Munyati. 2017b. "The Parasitic Weeds Scourge in Northern Zimbabwe: Effects of Soil Degradation, Hosts and Food Security Implications to Rural Farmers." *Scientia Agriculturae* 20: 3. doi:10.15192/pscp.sa.2017.20.3.8691.

Mbatudde, M., M. Mwanjololo, E. K. Kakudidi, and H. Dalitz. 2012. "Modelling the Potential Distribution of Endangered *Prunus Africana* (Hook. F .) Kalkm. In East Africa." *African Journal of Ecology* 50: 393–403. doi:10.1111/j.1365-2028.2012.01327.x.

Midega, C. A. O., C. J. Wasonga, A. M. Hooper, J. A. Pickett, and Z. R. Khan. 2017. "Drought-Tolerant Desmodium Species Effectively Suppress Parasitic Striga Weed and Improve Cereal Grain Yields in Western Kenya." *Crop Protection* 98: 94–101. doi:10.1016/j.cropro.2017.03.018.

Mohammadi, S., E. Ebrahimi, M. Shahriari Moghadam, and L. Bosso. 2019. "Modelling Current and Future Potential Distributions of Two Desert Jerboas under Climate Change in Iran." *Ecological Informatics* 52: 7–13. doi:10.1016/j.ecoinf.2019.04.003.

Mosomtai, G., M. Evander, P. Sandström, C. Ahlm, R. Sang, O. A. Hassan, H. Affognon, and T. Landmann. 2016. "Association of Ecological Factors with Rift Valley Fever Occurrence and Mapping of Risk Zones in Kenya." *International Journal of Infectious Diseases* 46. International Society for Infectious Diseases: 49–55. doi:10.1016/j.ijid.2016.03.013.

Mudereri, B. T., C. Mukanga, E. T. Mupfiga, C. Gwatirisa, E. Kimathi, and T. Chitata. 2020b. "Analysis of Potentially Suitable Habitat within Migration Connections of an Intra-African Migrant-the Blue Swallow (*Hirundo Atrocaerulea*)." *Ecological Informatics* 57: 101082. doi:10.1016/j.ecoinf.2020.101082.

Mudereri, B. T., T. Chitata, C. Mukanga, E. T. Mupfiga, C. Gwatirisa, and T. Dube. 2019b. "Can Biophysical Parameters Derived from Sentinel-2 Space-Borne Sensor Improve Land Cover Characterisation in Semi-Arid Regions? Space-Borne Sensor Improve Land Cover Characterisation." *Geocarto International*. 1–20. doi:10.1080/10106049.2019.1695956.

Mudereri, B. T., T. Dube, E. M. Adel-Rahman, S. Niassy, E. Kimathi, Z. Khan, and T. Landmann. 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. doi:10.5194/isprs-archives-XLII-2-W13-701-2019. June.

Mudereri, B. T., T. Dube, S. Niassy, E. Kimathi, T. Landmann, Z. Khan, and E. M. Abdel-Rahman. 2020a. "Is It Possible to Discern Striga Weed (Striga Hermonthica) Infestation Levels in Maize Agro-Ecological Systems Using in-Situ Spectroscopy?" *International Journal of Applied Earth Observation and Geoinformation* 85: 102008. doi:10.1016/j.jag.2019.102008.

Mugandani, R., M. Wuta, A. Makarau, and B. Chipindu. 2012. "RE-Classification of Agro-Ecological Regions of Zimbabwe in Conformity with Climate Variability and Change." *African Crop Science Journal* 20 (2): 361–369. https://pdfs.semanticscholar.org/d24a/4e02f534cf8e9f8705690c7a5ba19145becc.pdf

Muposhi, V. K., E. Gandiwa, A. Chemura, P. Bartels, S. M. Makuza, and T. H. Madiri. 2016. "Habitat Heterogeneity Variably

Influences Habitat Selection by Wild Herbivores in a Semi-Arid Tropical Savanna Ecosystem." *PloS One* 11: 9. doi:10.1371/journal.pone.0163084.

Naimi, B., and M. B. Araújo. 2016. "Sdm: A Reproducible and Extensible R Platform for Species Distribution Modelling." *Ecography* 39 (4): 368–375. doi:10.1111/ecog.01881.

Naimi, B., N. A. S. Hamm, T. A. Groen, A. K. Skidmore, and A. G. Toxopeus. 2014. "Where Is Positional Uncertainty a Problem for Species Distribution Modelling?" *Ecography* 37 (2): 191–203. doi:10.1111/j.1600-0587.2013.00205.x.

Nelder, J. A., and R. W. M. Wedderburn. 1972. "Generalized Linear Models." *Journal of the Royal Statistical Society. Series A: General* 135 (3): 370–384. doi:10.2307/2344614.

Niang, I., O. C. Ruppel, M. A. Abdrabo, A. Essel, C. Lennard, J. Padgham, and P. Urquhart. 2014. "Africa." In *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change Barros*, edited by L. L. White, V. R. Barros, C. B. Field, D. J. Dokken, M. D. Mastrandrea, K. J. Mach, T. E. Bilir, et al., 1199–1265. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. doi:10.1017/CBO9781107415386.002.

Ochungo, P., R. Veldtman, E. M. Abdel-Rahman, S. Raina, E. Muli, and T. Landmann. 2019. "Multi-Sensor Mapping of Honey Bee Habitats and Fragmentation in Agro-Ecological Landscapes in Eastern Kenya." *Geocarto International*. 1–22. doi:10.1080/10106049.2019.1629645.

Oswald, A., J. K. Ransom, J. Kroschel, and J. Sauerborn. 2001. "Transplanting Maize and Sorghum Reduces Striga Hermonthica Damage." *Weed Science* 49 (3): 346–353. doi:10.1614/0043-1745(2001)049[0346:TMASRS]2.0.CO;2.

Parker, C. 2009. "Observations on the Current Status of Orobanche and Striga Problems Worldwide." *Pest Management Science* 65 (5): 453–459. doi:10.1002/ps.1713.

Pickett, J. A., C. M. Woodcock, C. A. O. Midega, and Z. R. Khan. 2014. "Push-Pull Farming Systems." *Current Opinion in Biotechnology* 26: 125–132. doi:10.1016/j.copbio.2013.12.006.

Plant, R. E. 2012. *Spatial Data Analysis in Ecology and Agriculture Using R*. Califonia: CRC Press: Taylor and Francis Group.

QGIS Development Team. 2019. "QGIS Geographic Information System. Open Source Geospatial Foundation Project." http://qgis.osgeo.org

R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing,Vienna, Austria. https://Www.R-Project.Org/

Rich, P. J., and G. Ejeta. 2007. "Biology of Host-Parasite Interactions in Striga Species." In *Intergrating New Technologies for Striga Control: Towards Ending the Witch-Hunt* edited by Gebisa Ejeta and Jonathan Gressel, 19–32. Singapore: World Scientific Publishing Company.

Sasson, A. 2012. "Food Security for Africa: An Urgent Global Challenge." *Agriculture and Food Security* 1 (1): 1–16. doi:10.1186/2048-7010-1-2.

Spallek, T., M. Musembi, and K. Shirasu. 2013. "The Genus Striga: A Witch Profile." *Molecular Plant Pathology* 14 (9): 861–869. doi:10.1111/mpp.12058.

Stockwell, D. R. B., and A. T. Peterson. 2002. "Effects of Sample Size on Accuracy of Species Distribution Models." *Ecological Modelling* 148 (1): 1–13. doi:10.1016/S0304-3800(01)00388-X.

Sungirai, M., D. Z. Moyo, P. De Clercq, M. Madder, S. O. Vanwambeke, and E. M. De Clercq. 2018. "Modelling the Distribution of Rhipicephalus Microplus and R. Decoloratus in Zimbabwe." *Veterinary Parasitology: Regional Studies and Reports* 14: 41–49. doi:10.1016/j.vprsr.2018.08.006.

Tesfamichael, S. G., S. W. Newete, E. Adam, and B. Dubula. 2018. "Field Spectroradiometer and Simulated Multispectral Bands for Discriminating Invasive Species from Morphologically Similar Cohabitant Plants." *GIScience and Remote Sensing* 55 (3). Taylor & Francis: 417–436. doi:10.1080/15481603.2017.1396658.

Thamaga, K. H., and T. Dube. 2019. "Understanding Seasonal Dynamics of Invasive Water Hyacinth (Eichhornia Crassipes) in the Greater Letaba River System Using Sentinel-2 Satellite Data." *GIScience and Remote Sensing* 56 (8): 1355–1377. doi:10.1080/15481603.2019.1646988.

Vapnik, V. 1979. "Estimation of Dependences Based on Empirical Data." *Nauka, Moscow, Translation Springer Verlag, New York, 1982*. 27: 5165–5184.

Wan, J. Z., and C. J. Wang. 2019. "Determining Key Monitoring Areas for the 10 Most Important Weed Species under a Changing Climate." *Science of the Total Environment* 683: 568–577. doi:10.1016/j.scitotenv.2019.05.175.

Wan, Z., S. Hook, and G. Hulley. 2015. "MOD11C2 MODIS/Terra Land Surface Temperature/Emissivity 8-Day L3 Global 0.05Deg CMG V006 [Data Set]." *NASA EOSDIS Land Processes DAAC*. doi:10.5067/MODIS/MOD11C2.006.

Yoneyama, K., K. Yoneyama, Y. Takeuchi, and H. Sekimoto. 2007. "Phosphorus Deficiency in Red Clover Promotes Exudation of Orobanchol, the Signal for Mycorrhizal Symbionts and Germination Stimulant for Root Parasites." *Planta* 225 (4): 1031–1038. doi:10.1007/s00425-006-0410-1.