ISPRS Journal of Photogrammetry and Remote Sensing 159 (2020) 1–12

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

Wide-area invasive species propagation mapping is possible using phenometric trends

Tobias Landmann[a,b,*], Olena Dubovyk[c], Gohar Ghazaryan[c], Jackson Kimani[a], Elfatih M. Abdel-Rahman[a,d,*]

[a] International Centre of Insect Physiology and Ecology (icipe), P.O. Box 30772, 00100 Nairobi, Kenya
[b] Remote Sensing Solutions GmbH, Dingolfinger Str. 9, 81673 Munich, Germany
[c] Center for Remote Sensing of Land Surfaces (ZFL), University of Bonn, Walter-Flex-Str. 3, 53113 Bonn, Germany
[d] Department of Agronomy, Faculty of Agriculture, University of Khartoum, Khartoum North 13314, Sudan

ARTICLE INFO

Keywords:
MODIS EVI
Phenometrics
Land degradation
Bush encroachment
Logistic regression

ABSTRACT

Invasive plant species in eastern Africa severely impede rangeland and cropland productivity with dire consequences for livelihoods of agro-pastoralist communities. We produced the first occurrence and spread map of invasive plant species (Prosopis: *Prosopis juliflora* and Parthenium: *Parthenium hysterophorus*) for western Somaliland (a region of eastern Africa) using vegetation productivity and phenology trends from 250 m MODIS (Moderate Resolution Imaging Spectroradiometer) EVI (Enhanced Vegetation Index) time-series data (2001–2014). Binomial logistic regression models were created to predict the presence or absence of the invasive species from the MODIS EVI phenometrics and vegetation productivity trends. Model training pixels were extracted from a 30 m Landsat-based classification that mapped areas of propagation of the two invasive species between 2001 and 2015. Field observations collected during 2014 and 2015 were used as reference data for the Landsat classification. After optimization of the logistic regression models, a probability of occurrence map was produced and evaluated for each of the two invasive species. The probability maps predicted that the cropland-dominated areas in the southwestern part of Somaliland were considerably infested with Parthenium while Prosopis was most abundant in the peri-urban zones and the central and eastern regions. Vegetation amplitude (the seasonal cycle of vegetation between the vegetation peak and the trough) was most relevant and statistically significant for predicting the spread of Parthenium. This highlights the importance of vegetation seasonality variables for the wide-area mapping of herbaceous life forms in semi-arid biomes. Mann-Kendall trends based on annual summed EVI value and seasonal EVI peak value trends were the most relevant predictors for the occurrence of Prosopis. Phenometric trends show immense potential to map shifts in vegetation patterns in relation to the spread of invasive species as a consequence of global change effects, particularly in African drylands.

1. Introduction

There is growing global interest and need to study propagation zones of invasive species in order to support containment and comprehensive land restoration efforts (Fatimah and Ahmad, 2009; Singh et al., 2018). Despite the detrimental effects of invasive plant species on ecosystem services (especially on biodiversity, see Charles and Dukes, 2017) and agricultural productivity, a few studies on modelling the invasive plant species occurrence and spread have been conducted in tropical Africa (e.g. McConnachie et al., 2010).

Since invasive plants tend to spread rapidly across several biomes, wide-area assessment of their diversity, abundance and spread pathways is challenging (Huang and Asner, 2009). Mapping accuracy is affected by the difficulty of discriminating invasive species from co-occurring native species due to the similarity of their phenological responses to rainfall and temperature (Cleland et al., 2007). Subtle phenological responses such as flowering or abrupt greening may also preclude accurate discrimination in image reflectance data, especially when using satellite data with low temporal resolutions and limited spatial coverage (Wallace et al., 2016).

In some regions, for example the eastern African region of Somaliland (the focus region of the present study), the use of pure ecological niche modeling approaches for invasive species spread mapping is not always appropriate. This is due to the absence of

* Corresponding authors at: International Centre of Insect Physiology and Ecology (icipe), P.O. Box 30772, 00100 Nairobi, Kenya.
E-mail addresses: landmann@rssgmbh.de (T. Landmann), eabdel-rahman@icipe.org (E.M. Abdel-Rahman).

https://doi.org/10.1016/j.isprsjprs.2019.10.016
Received 17 May 2018; Received in revised form 23 October 2019; Accepted 24 October 2019
Available online 08 November 2019

historical occurrence data required for change modeling and because localized or small-scale land use factors (such as ecological disturbances from overgrazing or intensive cropping; Obiri, 2011) are considered to be better descriptors of invasive plant propagation than general climate zones and soil types (Early et al., 2016). As pure ecological niche modeling approaches may overestimate species occurrences because they do not consider actual land use and are biased towards climate predictors (Makori et al., 2017), methods integrating remote sensing may be more appropriate.

Land surface phenology, the seasonal and re-occurring variation of vegetation productivity over the land surface, can be estimated from remote sensing time-series data. (Zhang et al., 2006). Such phenometrics include all vegetation parameters that can be estimated from seasonal NDVI- (Normalized Differenced Vegetation Index) or EVI- (Enhanced Vegetation Index) based vegetation curves. Changes in phenology and vegetation productivity (as annually summed greenness) can be predicted using linear or Mann-Kendall trends. (Vrieling et al., 2013). In some instances, linear regression modeling, time-series smoothing functions using phenological markers or change vector amplitudes have been applied to time dependent NDVI or EVI.

The 250 m MODIS (Moderate Resolution Imaging Spectroradiometer) remotely-sensed time-series data sets (used as input data in the present study) are available for a period of over 15 years (2002 to present) and are better suited than bi-temporal data (i.e. multi-temporal Landsat satellite data) for land surface change characterization (Dubovyk et al., 2013). Also, while coarse resolution time-series data (> 500 m pixel resolution) on summed greenness cannot feasibly be linked to small-scale changes in phenological events and plant species composition (Julien and Sobrino, 2009), the use of phenometrics from medium resolution time-series data (< 250 m) shows great potential for mapping large-scale vegetation change in drylands, even under various land use conditions (Parplies et al., 2016). The use of the vegetation index means that methods can be applied to other regions and periods due to the availability of spectral wavebands for the vegetation index on most satellite instruments.

In Somaliland, Ng et al. (2016) used multi-temporal Landsat observations and field data to map the distribution of Prosopis (*Prosopis juliflora*) around the city Hargeysa for the year 2015 using a random forest classifier. However, no studies of eastern African drylands have utilized trends of phenometrics at a moderate resolution scale and logistic regression modeling to map the wide-area occurrence and propagation of invasive plant species. Studies of North American landscapes have used moderate resolution phenometrics and ancillary field data (density and distribution) to exploit differences in phenology between co-existing plants to map invasive grassland species (Bradley and Mustard, 2008; Huang and Geiger, 2008; Peterson, 2005; Wallace et al., 2016). Huang and Geiger (2008), used temporal segmentation and vegetation amplitudes from 250 m MODIS EVI to differentiate *E. lehmanniana* from native grassland species. They exploited the fact that natural grasses tended to grow less vigorously in wet and cooler periods. For a grassland biome, West et al. (2017) integrated seasonal vegetation vigor observations from 30 m Landsat with ecological covariates (such as slope, microclimate and fire regime) to model areas that are prone to the invasion of *Bromus tectorum* (cheatgrass). Where invasion occurs at a sub-canopy level and/or the landscape is more fragmented and complex (such as within or along urban areas in North America), Light detection and ranging (LiDAR)–based data sets have been successfully fused with very high-resolution (< 1 m pixel resolution) satellite (IKONOS) or airborne (AISA Eagle) imagery to accurately discriminate invasive vegetation (Singh et al., 2015; Peerbhay et al., 2016). Also, Singh et al. (2018) utilized multi-seasonal Landsat TM (Thematic Mapper) data to detect vegetation phenological responses in order to accurately map the distribution of understory plant invasion in urban forest ecosystem. In another study, Tarantino et al. (2019) employed a hybrid classification approach to map *Ailanthus altissima* invasion using 2 m WorldView-2 multi-temporal data. While the use of single or multi-season hyperspectral remote sensing provided accurate results for the localized mapping of invasive plants within various habitats in North America (He et al., 2011), United Kingdom (Taylor et al., 2013) and Brazil (Amaral et al., 2015), hyperspectral data sets are constrained to certain narrow phenological time windows and small detection footprints (Huang and Asner, 2009). Other studies have explored the use of unmanned aerial vehicles (UAV) with ultrahigh spatial resolution for rapid mapping and monitoring of invasive plants (e.g. Mafanya et al., 2017; Perroy et al., 2017).

The studies discussed all based their mapping or modeling approaches on the physiological and phenotypic characteristics of the invasive species, and all used an integrative approach, however their results are mostly restricted to single species, landscapes and/or time periods (i.e. years). Thus, the main objective of the present study was to show the usefulness of phenometrics and phenology profiles (predictors from moderate resolution satellite data) for estimating the occurrence and spread of two major (i.e. destructive and widespread) invasive plant species in Somaliland. Our research approach and main objective follows from the wide and fast spread of invasive species throughout the region (their eco-physiology) and the known potential of moderate resolution phenometrics trends (changes in annually summed vegetation productivity and phenology) to accurately map subtle vegetation shifts in semi-arid regions (Broich et al., 2014). A binomial (binary) logistic regression modeling approach (Agresti, 1996) was used in this study to model both invasive species. The binomial model permits the computation of accurate and comparable estimation coefficients. There is general consensus that more sophisticated statistical modeling routines that ideally consider biogeochemical traits at the landscape (i.e. localized) level could improve the prediction of phenological change from time-series data as a result of changes in species composition over time (Bradley and Mustard, 2008; Huang and Asner, 2009). Our study design further considers that we were able to collect both presence and absence point data for the two species in 2014/2015 and used 30 m Landsat observations as intermediate data sets to 'scale' the point data to the 250 m MODIS phenometrics data (Fisher and Mustard, 2007).

2. Methodology

2.1. Study area

The 8400 km^2 study area (indicated in Fig. 1 by the dashed rectangle) covers the peri-urban zone around Hargeysa, the capital city of the Republic of Somaliland, and is part of the central plateau region (Haud Plateau, indicated by the shaded area in Fig. 1) which exhibits distinct soil, elevation and vegetation landscape characteristics. The altitude of the study area ranges from 1200 to around 2100 m above mean sea level, and the land form is slightly undulating, almost flat and cut by several streams, with silty loam and clayey soils intermixed (Omuto and Vargas, 2009). The dominant land covers include open shrublands, sparse dense trees, herbaceous cover, urban and rural clusters, bare soils and waterbodies. The woody component of the study area is mainly sparsely distributed *Acacia* spp., underlain by palatable grasses (i.e. *Cenchrus ciliaris, Cynodon dactylon, Sporobolus marginatus, Tragus racemosus* and *Aristida adscensionis*) (Hadden, 2007).

The main weather patterns are determined by the passage of the seasonal monsoon winds. Somaliland thus has four seasons in the year: Gu (Spring, April to June; first wet season), Hagaa (Summer, July to August; short dry season), Dayr (Autumn, September to November; second wet season), and Jiilaal (Winter, December to March; long dry season). Although the semi-arid study region receives an average of 300–500 mm of annual rainfall, it is considerably variable and rain occurs erratically (Greenwood, 1957). This results in diverse inter-annual rainfall patterns (i.e. unimodal areas often become bimodal and vice versa).

The dominant land uses in the study region are wood collection and livestock grazing for milk production (mostly by nomadic herders).

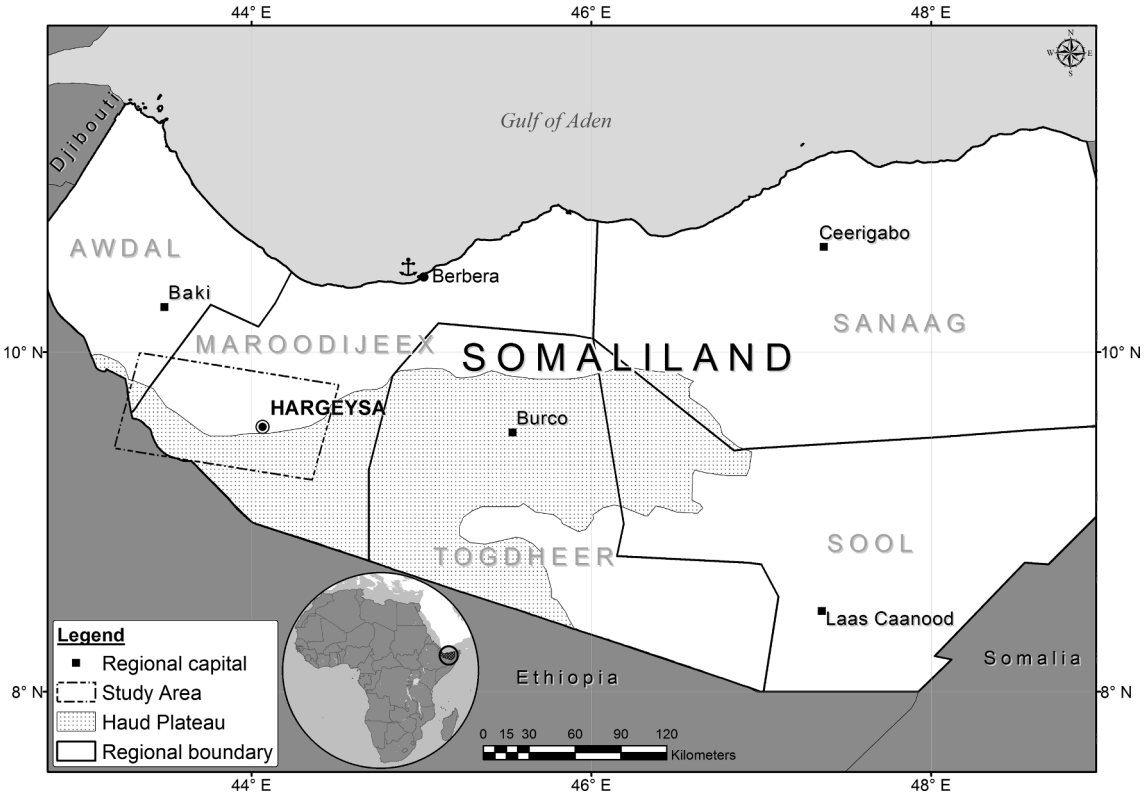

Fig. 1. The dashed rectangle shows the study area within Somaliland. Major regional boundaries are shown as black lines.

However, the southern and western sections are dominated by rain-fed (small-scale) cropping of maize and sorghum (Vargas et al., 2009).

2.2. Characteristics of the invasive species and their effects

The simultaneous invasion of croplands and rangelands by Parthenium and Prosopis has led to severe land degradation and biodiversity loss in Somaliland and the arid and semi-arid land regions in eastern Africa (Seta et al., 2013; Tabu et al., 2013). Prosopis is an evergreen shrub that often occurs as a mono-specific thicket along seasonal water courses and in relatively isolated and open rangelands with low vegetation densities (Berhanu and Tesfaye, 2006). Parthenium is an exceedingly invasive annual herbaceous weed that replaces indigenous fodder grasses in rangelands (Nigatu et al., 2010) and inhibits nitrogen fixation in soils in croplands (Tamado and Milberg, 2000). Both invasive species are fast growing, already well established and known to invade large tracts of both productive and already degraded land (Shiferaw et al., 2004).

The detrimental effects of both species on land productivity are aggravated by the lack of regional strategies for their management and containment. In eastern Africa, invasive species propagation control efforts are severely under-funded, disintegrated and haphazardly performed (Obiri, 2011). Moreover, technical capacity to utilize predictive models that would help to assess invasive species spread and risk areas over national boundaries is lacking (Shitanda et al., 2014). Cost effective and wide-area monitoring routines for invasive species are urgently needed for eastern Africa so that coherent intervention strategies can be effectively implemented.

2.3. Processing of EVI time-series data

Fig. 2 gives an overview of the input data sets and methodology used in the study. The input data consisted of MODIS EVI time-series data from the MOD13Q1 product (collection 5) for the years 2001–2014. EVI was used rather than NDVI because of the relative insensitivity of EVI to background soil reflectance (Huete et al., 2002). The 250 m MOD13Q1 image data were provided as atmospherically corrected 16-day composites. To account for noise artifacts, the time-series imagery was smoothed using the Savitzky-Golay filter algorithm (Atzberger and Eilers, 2011).

2.4. Computation of vegetation phenometrics and trends

Using the TIMESAT software (Jönsson and Eklundh, 2004), the following EVI phenometric indicators were computed (for each year) from the filtered EVI time-series data using a single growing season (unimodal) setting (Fig. 3): 'EVI trend', 'amplitude', 'small integral', 'large integral' 'peak EVI day', 'peak EVI value', 'right derivative', and 'left derivative'.

Some of the above phenometric variables, e.g. amplitude, small and large integrals can be used as integrative indicators of net primary productivity or the total vegetation production of a pixel over a specific time period (Jönsson and Eklundh, 2004). In this study, the assumption is made that the biological invasion of non-native species within a certain area alters the per pixel vegetation seasonality pattern, which is measurable by a change in phenology or total vegetation productivity over time (Vilà et al., 2011). If intra-annual or annual summed seasonal vegetation growth for a given area or pixel changes due to changes in vegetation composition (i.e. plant invasion), the intra-annual vegetation spectral response over longer periods also progressively changes (Vrieling et al., 2013). In Fig. 4, the mean annual summed EVI change or inter-annual spectral change (as a trend) is illustrated for selected pixels in the study area representing natural vegetation (n = 30) (nat. veg.,) and Prosopis (n = 30). The reference pixels (areas affected by the expansion of Prosopis and natural vegetation sites with no assumed change or infestation) were determined from the field observations taken in 2013 and 2014. Fig. 4 clearly shows that the two linear EVI trends are different in magnitude and direction. The trend for natural vegetation is slightly negative, while the trend for Prosopis is positive. The maximum EVI for the Prosopis increases over time.

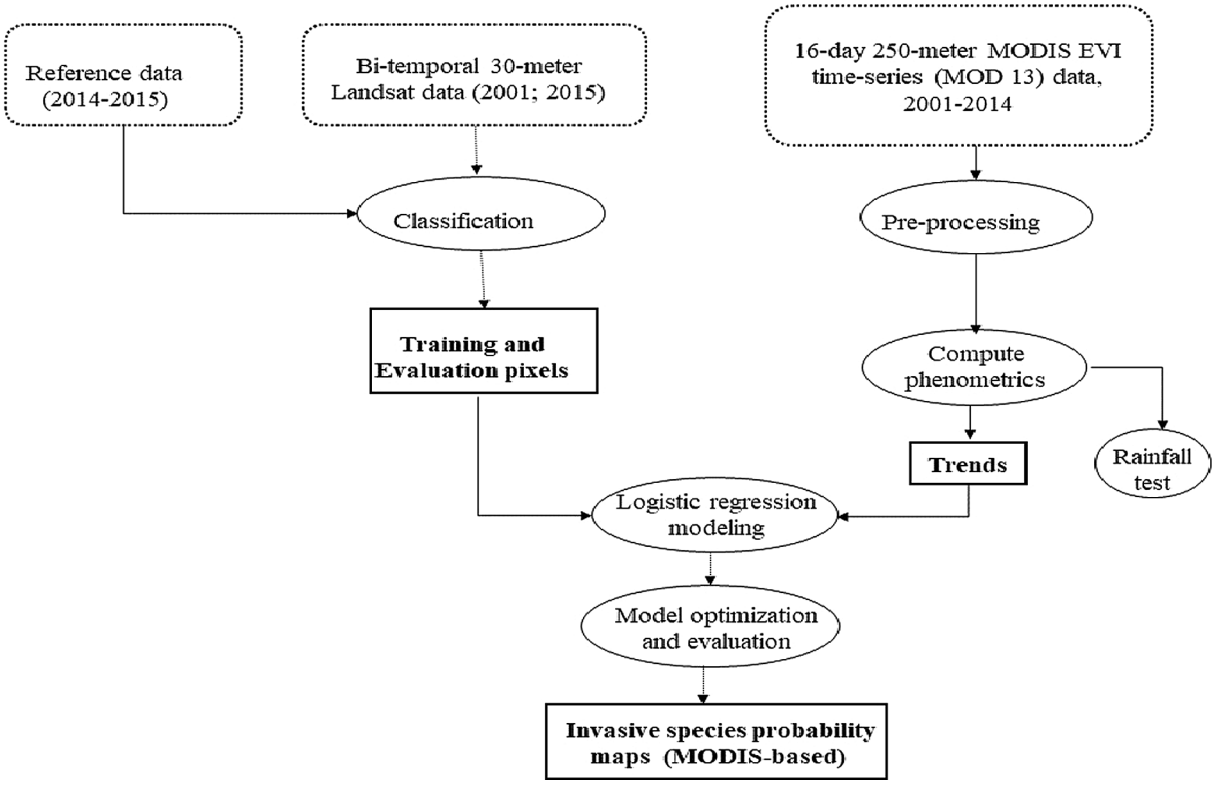

Fig. 2. Study methodology illustrating input data (dashed rectangles), primary data processing steps (ellipses) and derived or predicted data sets (rectangles). Arrows show data flow directions and tie-up points.

Because of the large intra-seasonal and inter-annual rainfall variability in drylands and in Somaliland in particular, phenometrics that utilize temporal markers pertaining to the start, end or length of the vegetation productivity seasons were not considered in this study (Bradley and Mustard, 2008).

A Mann-Kendall (MK) non-parametric trend analysis was applied to each of the phenometric variables to test for monotonic changes (as trends) over the observation period (2001–2014) (Mann, 1945). As opposed to linear trends, non-parametric MK trends provide significance levels for each trend while accounting for noise and seasonality effects in the time-series data. Mann Kendall trends were also computed using the annual summed EVI values ('EVI trend') (Fig. 4) (Neeti and Eastman, 2011).

A rainfall trend test (Fig. 2) using 1 km pixel resolution monthly

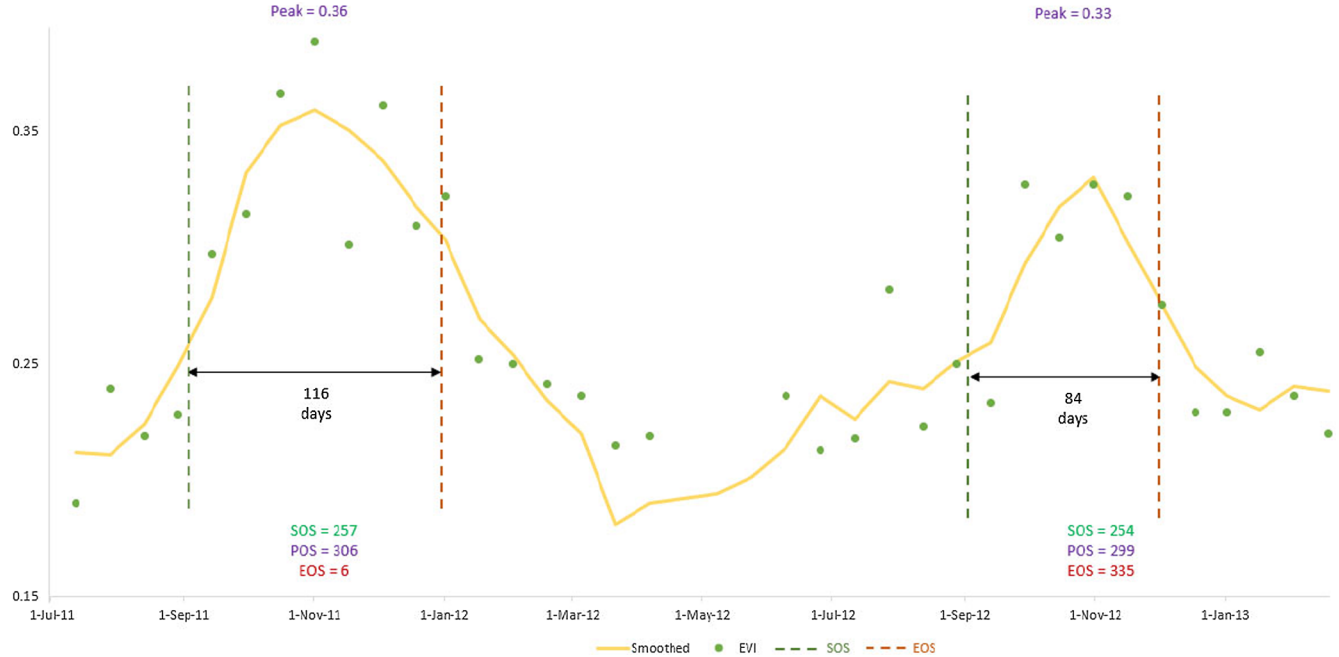

Fig. 3. Sample vegetation seasonality parameters for consecutive unimodal seasons generated in TIMESAT. EVI, SOS, POS and EOS are enhanced vegetation index, start of the season, peak of the season, and end of the season, respectively.

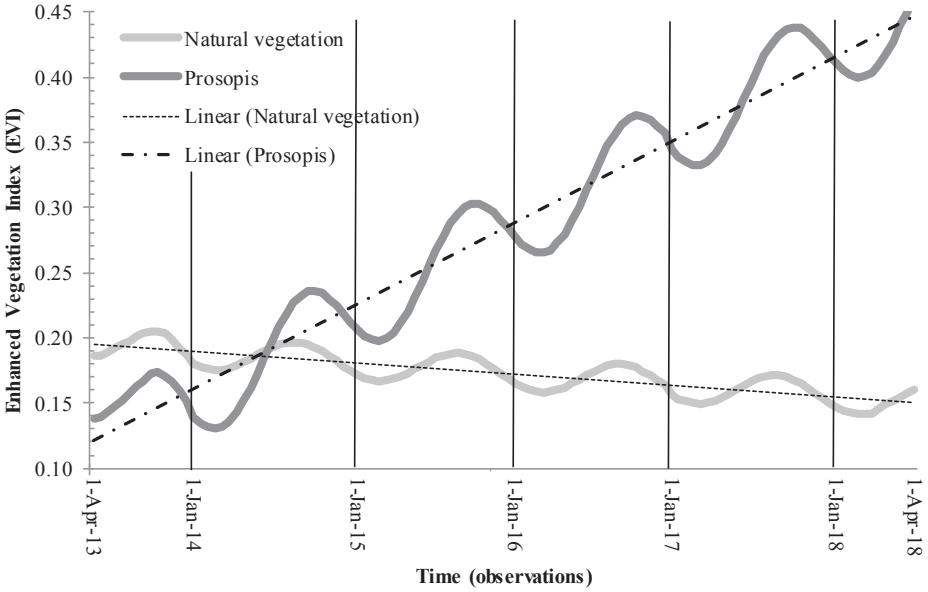

Fig. 4. EVI response curves for natural vegetation and Prosopis (*Prosopis juliflora*) estimated from harmonics corrected 30 Landsat enhanced vegetation index (EVI) data observations captured between April 2013 and April 2018. For illustrative purposes, best-fit harmonics corrections were applied to provide explicit trend lines.

rainfall grids from the AFRICLIM data set (Platts et al., 2015) indicated that rainfall could be excluded as a driver for vegetation change in the study area over the period 2001–2014.

2.5. Field data collection

Field reference data were collected using random sampling. Reference data ('Reference data' in Fig. 2) were collected for areas affected by the invasion of the two invasive species and for each of the following land cover/land cover change categories (land change classes): 'Acacia' (n = 144), 'Cropland' (n = 103), 'Natural Vegetation' (other than *Acacia* spp.; n = 63), 'Urban' (n = 163), 'Bare Soil' (n = 130), 'Deforestation' (n = 100), 'Parthenium' (n = 63), and 'Prosopis' (n = 91). The land change classes were determined by landscape composition and the most dominantly occurring classes, meaning that no underrepresented classes were sampled. At least 60 random samples for each class were collected. The maximum sample sizes for each class were determined by practical and logistical constraints largely due to the security situation in Somaliland. The sampling design was thus balanced according to the landscape composition but disproportional (with no homoscedascicity required) since no stratification was used (Glass et al., 1972).

Fig. 5 shows the spatial spread of the reference data. Reference points for 'Urban' and 'Bare soil' were collected through visual inspection of the Landsat image itself (these two classes are not shown in Fig. 5). Just like the field reference data, the visually determined reference data sets may also involve uncertainty; however, great care was taken to visually identify pixels that were clearly visible and temporally invariant.

Sampling was performed during the period from September 2014 to April 2015 to cover the seasonal cycle of natural vegetation and Parthenium, which is clearly visible in the vegetation growing season period from April to October. Prosopis is clearly visible throughout the year. A Global Positioning System (GPS) device with an accuracy of 3 m was used to tag the center point of each sampling site that had a size of at least 30 by 30 m. The following recordings were made for each sampling point: area name, date of sampling, site location (latitude and longitude), area size (m^2), percentage of trees/shrubs per sampling site (visually determined for 'Acacia' and 'Natural Vegetation') and homogeneity (also visually determined). Average homogeneity, defined as the fractional coverage (in %) of a land cover feature or class within pre-determined sampling areas, for 'Cropland', 'Parthenium' and 'Prosopis' was 95%. However, the number of trees/shrubs per sampling area/site for 'Acacia' and 'Natural Vegetation' varied considerably (with a mean of 20 and ranging between 4 and 44 trees/shrubs per area or site). The field records showed that greater than 89% of sites affected by non-native species had been invaded by the non-native species between 2000 and 2014/2015.

2.6. Landsat-based mapping of land cover change processes

Using a random forest classifier, the field reference data was used to produce a 30 m Landsat-based (explicit) classification map for the study area, including areas affected by the propagation of Prosopis and Parthenium between 2001 and 2015. The Landsat classification was then used to collect training and validation data points for logistic regression modeling (Section 2.7 below). In this work a 'study' concept is taken since associations between the MODIS-based phenological variables are investigated and class-specific comparisons were used for the Landsat mapping part. This omits the need for global i.e. class-specific proportion estimation for the reference data as well as the training and validation data (Glass et al., 1972).

Instead of collecting field reference data and scaling it to the 250 m MODIS-based phenometric trend data, the Landsat data were used as an intermediate data set to reduce scaling errors between the point-based field reference data and the relatively coarse MODIS data (Fisher and Mustard, 2007). Moreover, the Landsat mapping results (i.e. accuracy assessment using the reference data) were used to ascertain commission and omission errors in mapping of the two invasive species at moderate pixel resolutions in African drylands. Mapping the locations of invasive species would be possible with the spatially explicit bi-temporal Landsat data over a smaller area; however, the present study aimed to explore the potential of phenometric profiles (as trends) to map wide-area occurrences of the two invasive species.

The Landsat imagery data sets were resized to the spatial extent of the study area and the field data collections (Fig. 5) to reduce whole-scene spectral variability. MODIS-based phenometric profiles (i.e. their trends) were extracted as predictor variables for each of the Landsat mapped land change classes. The presence or absence of the two invasive species from the Landsat map were the response variables in the logistic regression models. The phenometrics corresponding to the presence of each of the two invasive species were coded as ones

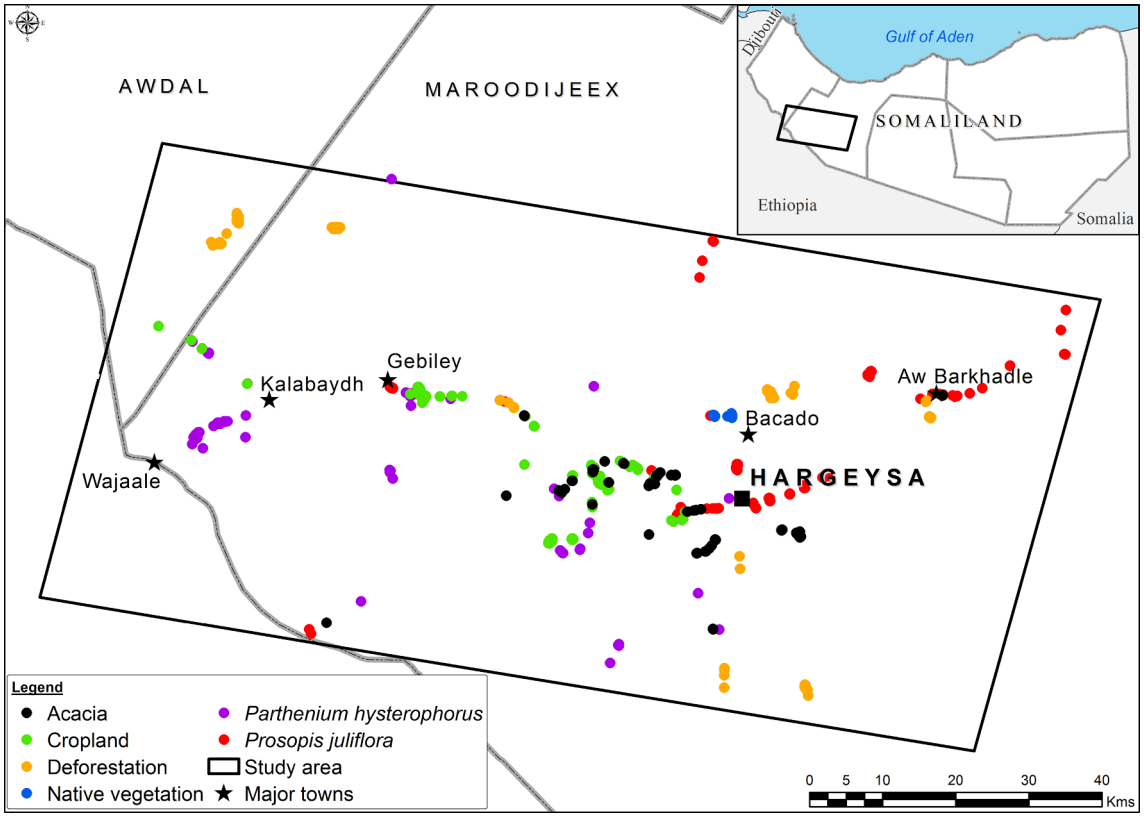

Fig. 5. Sampling sites (colored dots) of individual field (reference) data points belonging to each respective land change class collected randomly in the field during 2014 and 2015. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(presence), while the other land change classes were coded as zeros (absence).

Both bi-temporal Landsat image sets (Landsat 7 or the Enhanced Thematic Mapper Plus and Landsat 8 or the Operational Land Imager sensor) were acquired as surface reflectance corrected (Table 1). The two Landsat data sets use different surface reflectance correction algorithms (Masek et al., 2006). For each time period (2001 and 2015), EVI and the SAVI (Soil Adjusted Vegetation Index) data layers were computed and added to the reflectance bands.

A non-parametric random forest classifier was used to classify the bi-temporal and multi-seasonal Landsat layer stack ('Classification' in Fig. 2.). The random forest classifier is a machine learning classification tool that uses an ensemble of classification trees to produce a result. Each tree is built from a bootstrapped random sample containing approximately two-thirds of the training data drawn with replacement (Breiman, 2001). The remaining third of the reference data, i.e. the out-of-bag (OOB) samples, is used to internally evaluate the classification performance (Rodriguez-Galiano et al., 2012). The proportion of misclassifications of the total number of OOB elements essentially contributes an unbiased estimation of the generalization error (the OOB estimate of error rate). The OOB error for each class was used to create a confusion matrix for the eight classes. This confusion matrix was used to probe the accuracy of the Landsat mapped classes from which pixels were extracted as input for the logistic regression models. The random forest OOB error estimate is a reliable metric for assessing mapping accuracies, particularly when a relatively small number of field observations are available (e.g. we collected only 63 sample points for 'Parthenium') (Lawrence et al., 2006).

2.7. Selecting and extracting training pixels for logistic regression models

Model training and validation pixels that corresponded to the pixel values from the MODIS phenometric profiles for presence and absence of the two invasive species (according to the Landsat classification, response data) were selected and extracted for each land change class. The Landsat pixels were randomly and disproportionally selected around the class-specific field sampling points that were randomly selected in the field (Section 2.6). Essentially, pixels that were within a three by three Landsat pixel buffer area around the corresponding field data point were selected. If pixels from the entire Landsat map would have been selected and extracted, uncertainty errors would have been introduced due to re-sampling errors and class-specific Landsat

Table 1
Characteristics of the two bi-temporal Landsat data sets used for generating the Landsat-based classification (ETM + – Enhanced Thematic Mapper Plus, OLI – Operational Land Imager sensor, SR – Surface Reflectance).

Characteristics	Landsat-7 ETM +	Landsat-8 OLI
Processing level (original)	Level-1	Level-1
SR algorithms	(6S) radiative transfer model	Landsat Surface Reflectance Code (LaSRC) (internal algorithm)
Julian dates (season)	97 (dry); 305 (wet)	64 (dry); 288 (wet)
Acquisition years used	2001	2015
Mapping grid	Universal Transverse Mercator (UTM)	Universal Transverse Mercator (UTM)
SR bands used	Visible (1–5, 7) bands	Visible (1–7) bands

Table 2
Confusion matrix for the eight explicit land cover/land use and land cover change classes (land change classes) predicted from the bi-temporal Landsat data (Def – 'Deforestation', Nat. Veg – 'Natural vegetation', Parth – 'Parthenium', Pros – 'Prosopis').

Ground truth									
Predicted	Acacia	Cropland	Def.	Nat. Veg	Parth	Pros	Soil	Urban	User's accuracy (%)
Acacia	110	9	2	3	0	9	6	0	79.1
Cropland	1	120	6	4	3	1	0	0	88.9
Def	1	2	85	0	0	0	0	0	96.6
Nat. Veg	11	21	3	54	0	5	0	0	57.4
Parth	1	25	0	1	59	1	0	0	67.8
Pros	14	3	4	4	1	71	4	2	68.9
Soil	6	3	0	0	0	4	120	0	90.2
Urban	0	0	0	0	0	0	0	161	100.0
Total	**144**	**183**	**100**	**66**	**63**	**91**	**130**	**163**	
Producer's accuracy (%)	76.4	65.6	85.0	81.8	93.7	78.0	92.3	98.8	
Overall accuracy (%)	83								

classification inaccuracies over larger and spectrally inhomogeneous areas (Fisher, 1997). Spatial majority resampling (most common value) was used within an 8 by 8 pixel window to match the Landsat thematic data to the pixel resolution of the MODIS phenometrics data. The following total numbers of training and validation pixels were extracted from the re-sampled Landsat-based map: 90 for 'Parthenium', 194 for 'Prosopis', 347 for 'Deforestation', 146 for 'Croplands', 155 for 'Acacia' and 135 for 'Natural vegetation'. 70% of the total number of pixels for each class were used for training and the remaining 30% for validating the logistic regression model prediction results (Adelabu et al., 2015).

2.7.1. Logistic regression modeling

In logistic regression modeling, a linear relationship is generalized that relates a response variable to predictor variable(s) using a link function (McCullagh and Nelder, 1989). Before we fitted the logistic regression models, the phenometric trend variables were analyzed for co-linearity or multi-collinearity to improve model robustness (Fox, 2015). A factor analysis matrix (Tabachnick et al., 2001) was produced between each of the predictor variables, and a regression coefficient threshold of 0.7 was determined as a cut-off point. Using this coefficient threshold, no correlations between the phenometric trends (predictors) were found using the factor analysis matrix.

The use of a binomial logistic probability link function was chosen because of its ability to deal effectively with categorical variables and both normally and non-normally distributed predictor variables (e.g. data matrices of excessive smaller values or in some cases zeros as was the case in some of our data). The linear model is, in essence, the linear combination of a collection of predictor variables X_i to predict the response variable Y;

$$Y = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \cdots + \hat{\beta}_i X_i + \varepsilon \quad (1)$$

where $\hat{\beta}_0$ is the regression coefficient for the intercept and $\hat{\beta}_i$ represents the regression coefficient estimates for the explanatory variables 1 to i, computed between the response variable and the predictor variables X_i and ε is the error term.

The probability of occurrence for each phenometric variable (trend) in the logistic regression model is estimated as;

$$P(y = 1 \mid X_i) = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 x}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 x}} \quad (2)$$

where P is the probability of occurrence of an invasive species X_i between 0 and 1, and $\hat{\beta}_0$ and $\hat{\beta}_1$ are the regression coefficient estimates for each variable.

In our case, model coefficients were estimated for each invasive species class separately by setting the to-be-investigated invasive species class to 1 and the other classes to 0 in the independent (predictor) variable data column. An approach using separate logistic regression models for each invasive species was preferred over a single multinomial classification model since we aimed to produce separate accurate prediction probability maps for the two invasive plant species. Specifically, the prediction probability classes of both Parthenium and Prosopis were mapped using the natural breaks (Jenks) classification method (see Murray and Shyy, 2000), which assigns distinct break points to the values in the data set based on their natural grouping.

2.7.2. Model calibration and validation

The logistic regression models were initially calibrated using all phenometric trend variables for the Landsat-based training pixels or for only the field data. In a second iteration step, the models were optimized/re-calibrated (Fig. 2) using only the three most relevant (i.e. variables with the highest log odds ratios) and statistically significant ($p < 0.05$) variables contributing to the model fit (p-values and magnitudes of the regression coefficients) (Schwarz and Zimmermann, 2005). The p-values indicate the incompatibility between the data and the model used (Wasserstein and Lazar, 2016).

The optimized models were intrinsically evaluated using the accuracy of predicted fit and area under the curve (AUC) of Receiver Operating Curves (Fig. 2). The accuracy of the predicted data points is the percentage of the correctly predicted presence (coded as 1 in our model) as a result of the model fit (i.e. predicted fit); while the AUC measures if the probability of presence versus absence of a particular invasive species was correctly estimated by the logistic binary classification model (Powers, 2011). The AUC ranges between 0 (insignificant model) and 1 (model with good predictive ability).

3. Results

3.1. Spatially explicit Landsat mapping

The overall accuracy for the random forest classification result using all eight land change Landsat-based classes (including the invasive species classes) was 83% which can be considered an accurate result. The confusion matrix showed that 28% of the test pixels for 'Parthenium' were confused with pixels belonging to the class 'Cropland' (Table 2; User's accuracies), while there was also some confusion between the 'Prosopis' and 'Acacia' classes. The producer's accuracies showed good results for both species i.e. 78% and 91%, respectively. Given the relatively high user's, producer's and overall accuracies for the two invasive species classes, the Landsat-based training pixels and their corresponding phenometrics values could be used with confidence within the logistic regression model. Fig. 6 shows the importance of the Landsat bands stack (i.e. 2001 and 2015) for mapping the eight land change classes. The figure illustrates that blue and near infrared (NIR) bands were the most important predictor variables for mapping the land change classes in both the wet and dry seasons.

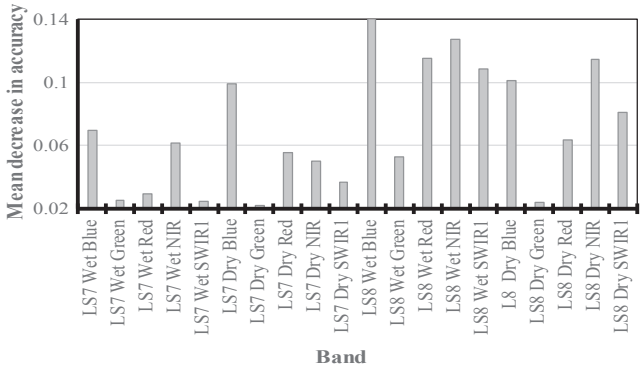

Fig. 6. The importance of Landsat bands for mapping the land cover/land cover change classes in the study area as measured using the random forest classification algorithm. LS7 and LS8 are Landsat 7 (in 2001) and Landsat 8 (in 2015), respectively. Wet and Dry are the wet and dry seasons during which the Landsat images were acquired. Blue, Green, Red, NIR and SWIR1 are the blue, green, red, near infrared and shortwave infrared 1 bands of LS7 and LS8.

3.2. Logistic regression model estimates

Table 3 shows the regression coefficient estimates (log odds, regression estimates, and *p*-values) for all phenometric trend variables in the two models.

In general, the regression log odds, the regression coefficient estimates ('Estimates' in Table 3 and 4) and the corresponding model fit significance levels (*p*-values) for the Parthenium model showed that the covariate 'amplitude' was statistically relevant for the propagation of Parthenium (log odds 262 and 533, $p < 0.001$ and a regression coefficient $> |5|$ using Landsat-based training pixels or only field data, respectively). This implied that 'amplitude' is a valid determinant to map the spread of Parthenium over the observation period (2001–2014). For the Prosopis models (i.e. for Landsat-based training pixels and only field data), the most relevant covariates were 'EVI trend' and 'Peak value' (for both covariates: log odds > 200, regression estimates $> |3|$ and $p < 0.001$).

3.3. Probability mapping using the phenometric trends

Fig. 7 (a, b) shows the probability results (maps) for the two optimized logistic regression models using the predicted probability of occurrences ($P(y = 1 \mid X)$ from (2)). The two occurrence probability maps in Fig. 7 (a, b) were optimized using only variables that were most relevant and significantly contributed to the model fit (log odds > 20, $p < 0.001$ and greater (absolute) regression coefficient values, respectively, for both models). The optimization step was performed to increase the explanatory power of the prediction estimates and essentially attain better prediction results.

For Parthenium, higher probability areas were mostly the cropland-dominated regions (see inset in Fig. 7a around the town of Kalabayadh). The grayscale probability image was classified into three classes based on the Jenks natural breaks algorithm; values in the range 0.01–0.10 were classified as low probability, values in the range 0.11–0.30 were classified as medium probability and values in the range 0.31–0.93 were classified as high probability. The Prosopis prediction results (Fig. 7b) showed similar patterns but there were also higher probability areas in peri-urban areas (e.g. Hargeysa). Values in the range 0.06–0.18 were classified as low probability, values in the range 0.19–0.56 were classified as medium probability and values in the range 0.57–0.99 were classified as high probability. There were also clear indications of severe infestation levels along major seepage lines and within riverside vegetation (see inset in Fig. 7b around the town of Aw Barkhadle). Prosopis was also most abundant in the rangeland areas in the eastern and central parts of the study area.

3.4. Model validation results

The reduced (optimized) model for Parthenium had a prediction accuracy of 0.89 (the prediction accuracy for the non-optimized model was 0.86) and an AUC score of 0.85 (AUC for the non-optimized model was 0.83), while for Prosopis the optimized and non-optimized model accuracy scores were slightly greater (prediction accuracy = 0.94 and 0.87; AUC = 0.93 and 0.89 for the optimized and non-optimized models, respectively). The acceptable prediction accuracies and AUC scores for both optimization models (Fig. 8) suggested a very high predictive accuracy for each modeled species. The above-mentioned accuracy scores were estimated using the validation pixels (samples) from the Landsat-based sampling (see Section 2.7). When only field data were used, the prediction accuracy for the optimized models was 0.93 for both Parthenium and Prosopis. While, the AUCs were 0.95 and 0.94 for Parthenium and Prosopis, respectively.

4. Discussion

4.1. Overall approach and suitability of variables

In this study, we aimed to show the potential of using phenometric profiles for predicting the occurrence of two major invasive plant species in African drylands. In particular, we investigated the use of 250 m MODIS time-series data and optimized logistic regression models for mapping the two invasive plants.

Overall, the relevance and statistical sensitivity of the phenometric trends (covariates in Tables 3 and 4) and the acceptable prediction accuracies of the two optimized models (prediction accuracies > 0.89 for both optimized models) highlights the suitability of phenometrics for invasive plants occurrence and propagation mapping in drylands. The suitability of phenometric trends from moderate resolution time-series data to map subtle biome shifts (also as a result of large-scale land

Table 3
Logistic regression model coefficient estimates for the eight MODIS-based phenometric trends (covariates) in the two models using Landsat-based training pixels.

Parthenium					Prosopis			
	log odds	Regression Estimates	p-value		log odds	Regression Estimates	p-value	
(Intercept)	0.04	−3.13	0	*	0.05	−3.02	0.	*
EVItrend	2.21	0.79	0.39	.	203.36	5.36	0.	*
amplitude	262.16	5.57	0	*	0.04	−1.54	0.16	.
sintegral	0.00	−6.25	0	*	41.32	3.53	0.03	*
lintegral	3.33	1.20	0.33	.	38.40	3.81	0.01	*
Peakday	17.94	2.89	0.02	*	53.93	3.88	0	*
Peakvalue	0.01	−4.31	0	*	462.72	6.03	0	*
rderivative	21.29	3.06	0	*	0.40	−0.93	0.38	.
lderivative	44.83	3.80	0	*	0.02	−4.13	0	*

Significance codes: < 0.05 '*' > 0.05 '.'

Table 4
Logistic regression model coefficient estimates for the eight MODIS-based phenometric trends (covariates) in the two models using only field data.

Parthenium					Prosopis			
	log odds	Regression Estimates	p-value		log odds	Regression Estimates	p-value	
(Intercept)	0.01	−4.46	0.00	*	0.04	−3.20	0.00	*
EVItrend	3.22	1.17	0.07	.	43.12	3.76	0.00	*
amplitude	533.79	6.28	0.03	*	0.02	−4.16	0.22	.
sintegral	0.89	−0.14	0.96	.	5.60	1.72	0.60	.
lintegral	40.04	3.69	0.00	*	83.42	4.42	0.07	.
Peakday	21.54	3.07	0.01	*	25.59	3.24	0.03	*
Peakvalue	0.00	−12.59	0.00	*	862.64	6.76	0.00	*
rderivative	609.10	6.41	0.00	*	0.38	−0.98	0.51	.
lderivative	7.69	2.04	0.56	.	0.02	−3.86	0.02	*

Significance codes: < 0.05 '*' > 0.05 '.'

management regimes) in heterogeneous semi-arid landscapes has been confirmed by several studies (Dubovyk et al., 2013; Senf et al., 2013; Wessels et al., 2012; Singh et al., 2018).

The positive regression coefficient estimates for vegetation amplitude in the Parthenium model (5.57 and 6.28; Tables 3 and 4, respectively) and the relatively large log odds values (262 and 533; Tables 3 and 4, respectively) highlight the relative importance of this variable for the spread of Parthenium in the study area. An increase in seasonal accumulated vegetation activity likely occurred due to the successive invasion of degraded sites alongside or within croplands by this annual weed (Zuberi et al., 2014). Vegetation amplitude, as a phenometric variable, is a good indicator for seasonal accumulated vegetation productivity, and if rainfall can be excluded as a vegetation productivity determinant over time, it is a good measure of vegetation productivity change for herbaceous vegetation communities (i.e. annual grasses, crops or weeds) in semi-arid landscapes (Mosomtai et al., 2016). The importance of mean seasonal annual EVI trend and increase in the largest EVI over time to Prosopis (Tables 3 and 4) was not surprising since substantial tree cover increases result in an increase in the cumulative vegetation productivity signal over time (annual greenness measurable by EVI or NDVI) (Runnström, 2000) (Fig. 5). It is worth noting that we used only three, yet relevant variables (selected according to their log odds ratios and p-value < 0.05) to generate the invasive species risk and propagation maps (Fig. 7). In general, models with a few predictor variables are more reliable, unbiased and do not overfit (Reunanen, 2003; Abdel-Rahman et al., 2013). Our models that developed using only three variables outperformed (AUC was 0.95 and 0.94 for Parthenium and Prosopis, respectively) models that generated using all the variables (AUC was 0.83 and 0.89 for Parthenium and Prosopis, respectively). This supports the suitability of our variable selection approach. Notwithstanding, our empirical approach of selecting only three variables, which essentially based on expert knowledge (Hurlbert, et al., 2019; Wasserstein et al., 2019), could be subjective, hence needs further investigation.

4.2. Statistical modeling and reference data

Binomial logistic regression modeling is seldom used in remote sensing studies. However, in some studies it has been found to be superior to currently established machine learning approaches, such as random forest, which tends to ignore error distributions such as asymmetry and heteroscedasticity in the response variables (e.g. count data) (Lopatin et al., 2016). In our case, the use of logistic regression modeling can be justified by the data characteristics and data distribution. Essentially, a higher number of absence than presence points were used in this study (Section 2.7), which mitigates model overestimation, while the data were non-normally distributed and the predictor variables were found to be non-collinear (Singh et al., 2015). By using model training and validation data with known accuracies from the 30 m Landsat data (Landsat and field based), this study aimed to reduce scaling effects in transferring point data to large pixel sizes (i.e. > 250 m) with inherent large spatial variabilities (Foody, 2002). For the Landsat mapping and comparative modeling, reference data were collected in the field using a probability-based (random) sampling design. Since we employed a study concept, and not a survey-adjusted approach with a need for global estimates, a sample subset of the whole study area would suffice, given permissible and known accuracies of the reference data. As such, we also did not consider error propagation per strata over the whole study area (Glass et al., 1972).

4.3. Comparable studies

There are no comparable studies that investigated the use of several phenometrics and their trends from moderate-resolution time-series data to map invasive species occurrence and propagation zones in eastern Africa. The modeling results in this study, however, confirmed what other localized field surveys found. Both species occur widely within the cropland areas in south-western Somaliland, near or within urban areas, in sites dominated by deep soils (such as along water courses) and in areas where rainfall is relatively abundant (Mills et al., 2015; Ng et al., 2016). The propagation of Parthenium in croplands can be largely explained by agriculture intensification (i.e. transport and spread of seeds through farming equipment, etc.). In regions in which the capacity to control or mitigate invasive plants is low, such as eastern Africa, agricultural intensification is a particularly important factor as a spread mechanism (Early et al., 2016).

Many studies in the Americas have shown the potential of individual phenological variables to predict invasive species infestation rates for specific landscapes by temporally assessing subtle differences in green-up responses between native and non-native vegetation communities (Huang and Geiger, 2008; Peterson, 2005; Wallace et al., 2016; Singh et al., 2018). In some studies, vegetation greening in moderate resolution data (30–250 m resolution) measured by NDVI or EVI, was linked to synoptic climate observations within unimodal rainfall distribution settings (Huang and Geiger, 2008; Wallace et al., 2016), with overall mapping accuracies ranging from 49% to 66%. However, in most cases the occurrence of only one invasive species within a certain landscape was investigated (Xie et al., 2008; Wallace et al., 2016; Peerbhay et al., 2016).

5. Conclusions

In this study, we assessed the applicability of MODIS-based phenometrics and logistic regression modeling for mapping propagation areas of two important invasive species in Somaliland. The results confirmed the potential of phenometric trends measurable from moderate resolution time-series data for wide-area assessment of invasive species in drylands given the integration of accurate contextual information (i.e. field data). The relevance and statistical contribution of the phenometric trends to the logistic regression model varied

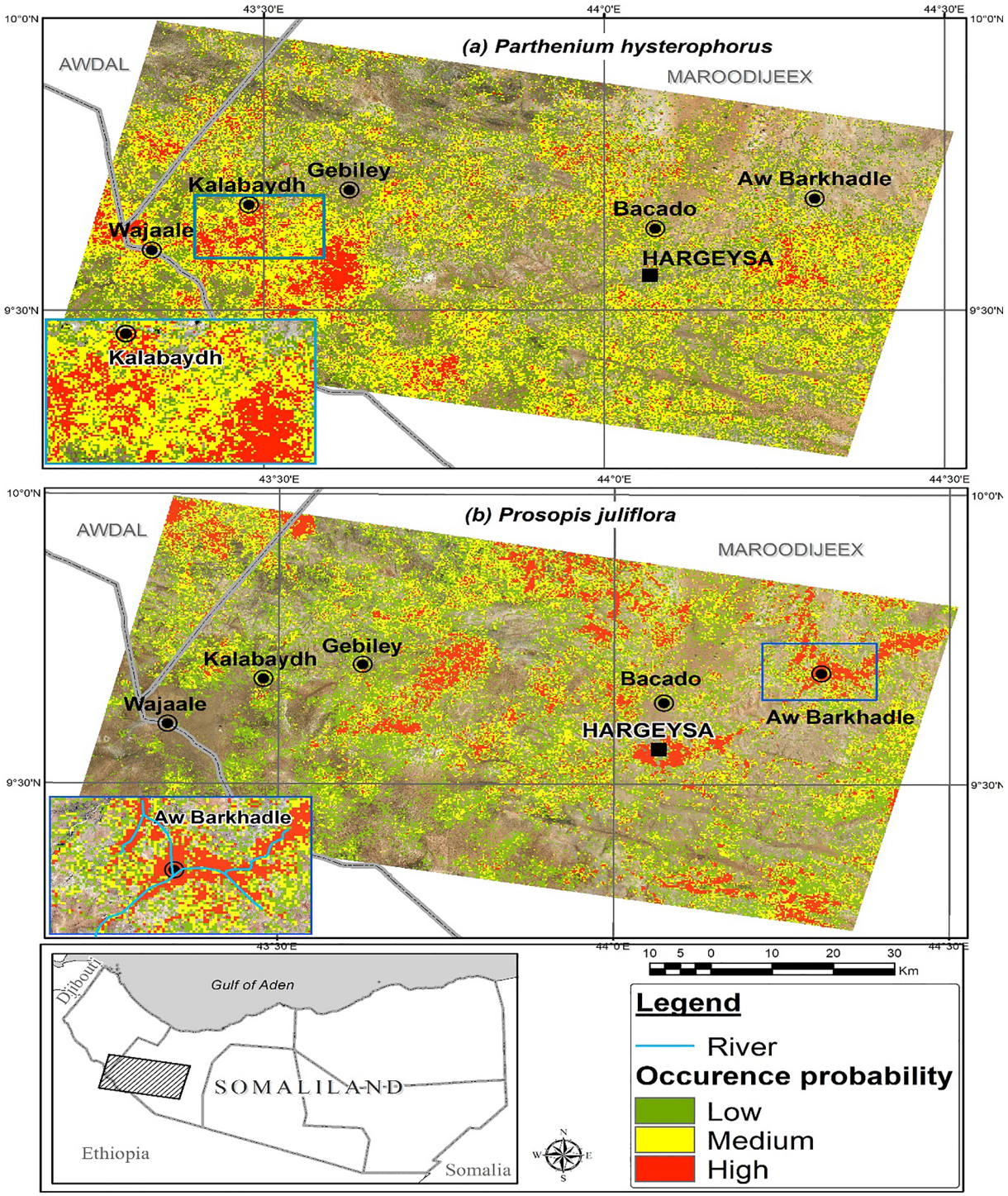

Fig. 7. Predicted probabilities for Parthenium (*Parthenium hysterophorus*) in (a) using the phenometric trend variables 'amplitude', 'right derivative' and 'left derivative', and in (b) for Prosopis (*Prosopis juliflora*) using the variables 'EVI trend', 'peak day' and 'peak value'. Red areas correspond to high occurrence probabilities, while the yellow-orange colors represent medium probabilities. The blue rectangles show the location of the insets. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

according to the invasive species being mapped: vegetation amplitude was the most relevant variable for Parthenium while MK trends based on annual summed EVI value and seasonal EVI peak value trends were the two most relevant for the propagation of Prosopis. Given the assimilation of well-processed covariates that are not auto-correlated and occurrence data sets that are binary or ordinal, the use of logistic regression modeling for mapping the probability of occurrence and propagation of non-native species, is highly recommendable.

Predictive maps like those presented here can be used to identify risk, buffer and containment zones in support of natural resources and invasive species management strategies. Likewise, they can be used to create awareness of the extent of invasive species that may not be currently well recognized as pertinent environmental issues but have dire consequences for biodiversity and agricultural productivity in many parts of Africa. With the advent of better available synergetic time-series observations from 30 m Landsat and 10–20 m Sentinel-2

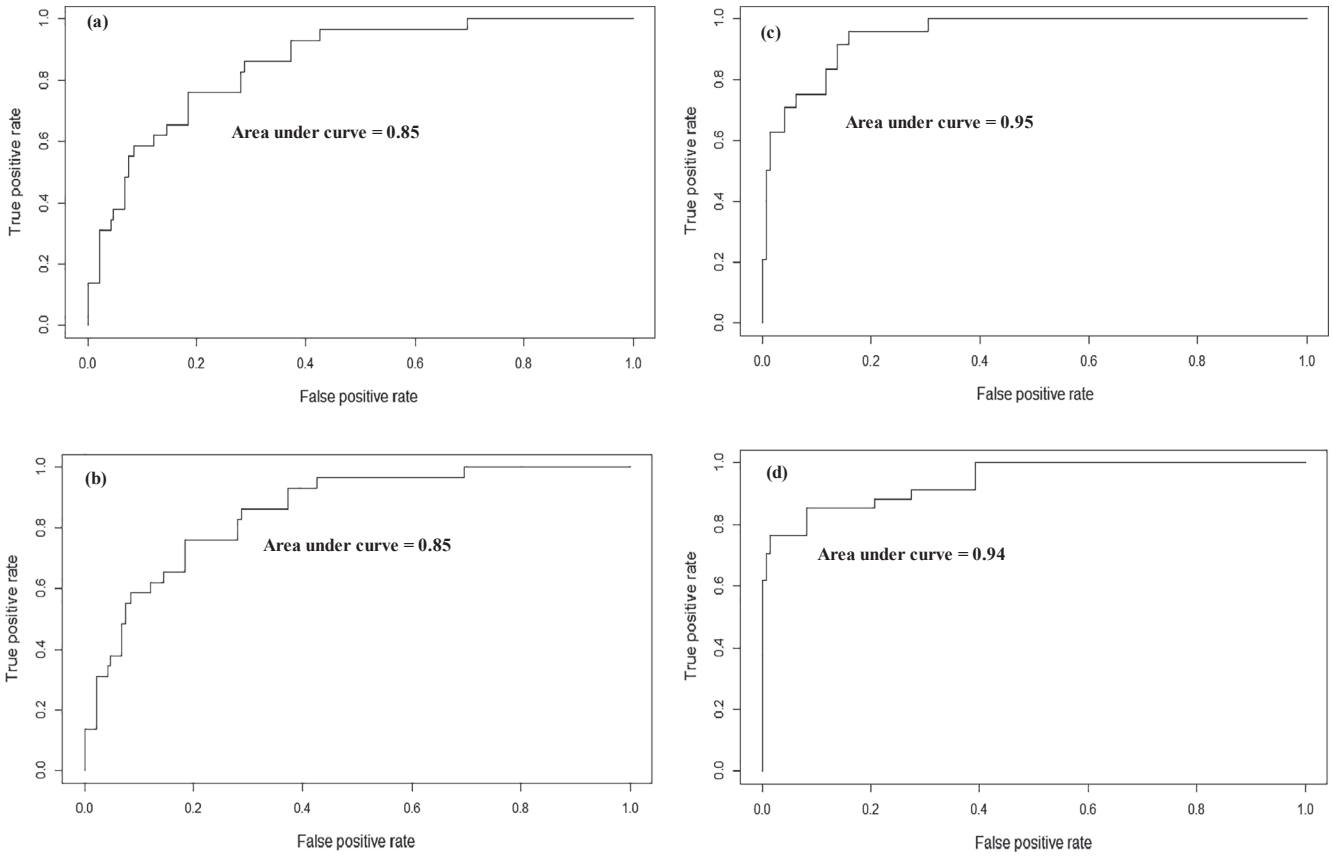

Fig. 8. Receiver Operating Curves for the logistic regression models: (a) predicting Parthenium using Landsat-based training pixels, (b) predicting Parthenium using only field data, (c) predicting Prosopis using Landsat-based training pixels, and (d) predicting Prosopis using only field data.

data streams, there is need to further investigate the use of spatially more explicit phenometrics to map subtle changes in vegetation dynamics and gradients in relation to invasive species propagation and global change effects.

Acknowledgements

We gratefully acknowledge the financial support for this research by the following organizations and agencies: The European Union, project FED/2013/319-933; the Centre for International Migration and Development (CIM) of the German Development Agency (GIZ), Germany; UK's Department for International Development (DFID), United Kingdom; Swedish International Development Cooperation Agency (Sida), Sweden; the Swiss Agency for Development and Cooperation (SDC), Switzerland; and the Kenyan Government, Kenya. The views expressed herein do not necessarily reflect the official opinion of the donors. Furthermore, we are grateful to Dr. Adan Elmi Abdullahi and his team for field data collections and the project team in Hargeysa, especially Dr. Mohamed Hassan, for supporting this action. Additionally, we would like to acknowledge the University of Bonn, Germany, and specifically the Center for Remote Sensing of Land Surfaces for their contribution in data processing. We are sincerely thankful to the Ministry of Livestock in Somaliland and the Food and Agricultural Organization of the UN for facilitating project implementation. Our utmost gratitude goes to Jandouwe Villinger for proof reading and other colleagues at *icipe* who helped in the technicalities of the classification algorithms, foremostly Richard Kyalo and Emily Kimathi. Lastly, we sincerely thank Dr Guido Lüchters for his excellent guidance on statistical modeling.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.isprsjprs.2019.10.016.

References

Abdel-Rahman, E.M., Ahmed, F.B., Ismail, R., 2013. Random forest regression and spectral band selection for estimating sugarcane leaf nitrogen concentration using EO-1 Hyperion hyperspectral data. Int. J. Remote Sens. 34, 712–728.
Adelabu, S., Mutanga, O., Adam, E., 2015. Testing the reliability and stability of the internal accuracy assessment of random forest for classifying tree defoliation levels using different validation methods. Geocarto Int. 30, 810–821.
Agresti, A., 1996. Categorical Data Analysis. John Wiley & Sons Inc., New York.
Amaral, C.H., Roberts, D.A., Almeida, T.I.R., Filho, C.R.S., 2015. Mapping invasive species and spectral mixture relationships with neotropical woody formations in southeastern Brazil. ISPRS J. Photogramm. Remote Sens. 108, 80–93.
Atzberger, C., Eilers, P.H.C., 2011. Evaluating the effectiveness of smoothing algorithms in the absence of ground reference measurements. Int. J. Remote Sens. 32, 3689–3709.
Berhanu, A., Tesfaye, G., 2006. The Prosopis dilemma, impacts on dryland biodiversity and some controlling methods. J. Drylands. 1, 158–164.
Bradley, B.A., Mustard, J.F., 2008. Comparison of phenology trends by land cover class: a case study in the Great Basin, USA. Glob. Change Biol. 14, 334–346.
Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.
Broich, M., Huete, A., Tulbure, M.G., Ma, X., Xin, Q., Paget, M., Restrepo-Coupe, N., Davies, K., Devadas, R., Held, A., 2014. Land surface phenological response to decadal climate variability across Australia using satellite remote sensing. Biogeosciences. 11, 5181–5198.
Charles, H., Dukes, J.S., 2017. Impacts of Invasive Species on Ecosystem Services. In: Nentwig, W. (Ed.), Ecological Studies. Springer, pp. 217–237.
Cleland, E.E., Chuine, I., Menzel, A., Mooney, H.A., Schwartz, M.D., 2007. Shifting plant phenology in response to global change. Trends Ecol. Evol. 22, 357–365.
Dubovyk, O., Menz, G., Conrad, C., Kan, E., Machwitz, M., Khamzina, A., 2013. Spatio-temporal analyses of cropland degradation in the irrigated lowlands of Uzbekistan using remote-sensing and logistic regression modeling. Environ. Monit. Assess. 185, 4775–4790.
Early, R., Bradley, B.A., Dukes, J.S., Lawler, J.J., Olden, J.D., Blumenthal, D.M., Gonzalez, P., Grosholz, E.D., Ibañez, I., Miller, L.P., Sorte, C.J.B., Tatem, A.J., 2016. Global threats from invasive alien species in the twenty-first century and national response capacities. Nat Commun. 7, 12485.

Fatimah, H., Ahmad, T., 2009. Phenology of Parthenium hysterophorus–a key factor for the success of its invasion. Adv. Environ. Biol. 3, 150–156.
Fisher, J.I., Mustard, J.F., 2007. Cross-scalar satellite phenology from ground, Landsat, and MODIS data. Remote Sens. Environ. 109, 261–273.
Fisher, P., 1997. The pixel: a snare and a delusion. Int. J. Remote Sens. 18, 679–685.
Foody, G.M., 2002. Status of land cover classification accuracy assessment. Remote Sens. Environ. 80, 185–201.
Fox, J., 2015. Applied Regression Analysis and Generalized Linear Models. SAGE Publications, Washington DC.
Glass, G.V., Peckham, P.D., Sanders, J.R., 1972. Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. Rev. Educ. Res. 42, 237–288.
Greenwood, J.E.G.W., 1957. The development of vegetation patterns in Somaliland Protectorate. Geogr. J. 123, 465–473.
Hadden, R.L., 2007. Geology of Somalia: A Selected Bibliography of Somalian Geology, Geography and Earth Science. Topographic Engineering Center, U.S. Army Corps of Engineers, Alexandria, Va.
He, K.S., Rocchini, D., Neteler, M., Nagendra, H., 2011. Benefits of hyperspectral remote sensing for tracking plant invasions. Divers. Distrib. 17, 381–392.
Huang, C., Asner, G.P., 2009. Applications of remote sensing to alien invasive plant studies. Sensors 9, 4869–4889.
Huang, C., Geiger, E.L., 2008. Climate anomalies provide opportunities for large-scale mapping of non-native plant abundance in desert grasslands. Divers. Distrib. 14, 875–884.
Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens. Environ. 83, 195–213.
Hurlbert, S.H., Levine, R.A., Utts, J., 2019. Coup de Grâce for a tough old bull: "statistically significant" expires. Am Stat. 73, 352–357.
Jönsson, P., Eklundh, L., 2004. TIMESAT—a program for analyzing time-series of satellite sensor data. Comput. Geosci. 30, 833–845.
Julien, Y., Sobrino, J.A., 2009. Global land surface phenology trends from GIMMS database. Int. J. Remote Sens. 30, 3495–3513.
Lawrence, R.L., Wood, S.D., Sheley, R.L., 2006. Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (randomForest). Remote Sens. Environ. 100, 356–362.
Lopatin, J., Dolos, K., Hernández, J., Galleguillos, M., Fassnacht, F.E., 2016. Comparing generalized linear models and random forest to model vascular plant species richness using LiDAR data in a natural forest in central Chile. Remote Sens. Environ. 173, 200–210.
Mafanya, M., Tsele, P., Botai, J., Manyama, P., Swart, B., Monate, T., 2017. Evaluating pixel and object based image classification techniques for mapping plant invasions from UAV derived aerial imagery: Harrisia pomanensis as a case study. ISPRS J. Photogramm. Remote Sens. 129, 1–11.
Makori, D., Fombong, A., Abdel-Rahman, E.M., Nkoba, K., Ongus, J., Irungu, J., Mosomtai, G., Makau, S., Mutanga, O., Odindi, J., Raina, S., Landmann, T., 2017. Predicting spatial distribution of key honeybee pests in Kenya using remotely sensed and bioclimatic variables: key honeybee pests distribution models. ISPRS Int. J. Geo-Inf. 6, 66.
Mann, H.B., 1945. Nonparametric tests against trend. Econometrica 13, 245–259.
Masek, J.G., Vermote, E.F., Saleous, N.E., Wolfe, R., Hall, F.G., Huemmrich, K.F., Gao, F., Kutler, J., Lim, T.-K., 2006. A Landsat surface reflectance dataset for North America, 1990–2000. IEEE Geosci. Remote Sens. Lett. 3, 68–72.
McConnachie, A.J., Strathie, L.W., Mersie, W., Gebrehiwot, L., Zewdie, K., Abdurehim, A., Abrha, B., Araya, T., Asaregew, F., Assefa, S., Gebre-Tsadik, R., Nigatu, L., Tadesse, B., Tana, T., 2010. Current and potential geographical distribution of the invasive plant Parthenium hysterophorus (Asteraceae) in eastern and southern Africa. Weed Res. 10, 1–14.
McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models, 2nd ed. CRC Press, New York.
Mills, M.S., Cohen, C., Francis, J., Spottiswoode, C.N., 2015. A survey for the critically endangered Liben Lark *Heteromirafra archeri* in Somaliland, north-western Somalia. Ostrich. 86, 291–294.
Mosomtai, G., Evander, M., Sandström, P., Ahlm, C., Sang, R., Hassan, O.A., Affognon, H., Landmann, T., 2016. Association of ecological factors with Rift Valley fever occurrence and mapping of risk zones in Kenya. Int. J. Infect. Dis. 46, 49–55.
Murray, A.T., Shyy, T.-K., 2000. Integrating attribute and space characteristics in choropleth display and spatial data mining. Int. J. Geogr. Inf. Sci. 14, 649–667.
Neeti, N., Eastman, J.R., 2011. A contextual Mann-Kendall approach for the assessment of trend significance in image time Series. Trans. GIS. 15, 599–611.
Ng, W.-T., Meroni, M., Immitzer, M., Böck, S., Leonardi, U., Rembold, F., Gadain, H., Atzberger, C., 2016. Mapping *Prosopis* spp. with Landsat 8 data in arid environments: Evaluating effectiveness of different methods and temporal imagery selection for Hargeisa, Somaliland. Int. J. Appl. Earth Obs. Geoinf. 53, 76–89.
Nigatu, L., Hassen, A., Sharma, J., Adkins, S.W., 2010. Impact of *Parthenium hysterophorus* on grazing land communities in north-eastern Ethiopia. Weed Biol. Manag. 10, 143–152.
Obiri, J.F., 2011. Invasive plant species and their disaster-effects in dry tropical forests and rangelands of Kenya and Tanzania. Jàmbá J. Disaster Risk Stud. 3, 417–428.
Omuto, C.T., Vargas, R.R., 2009. Combining pedometrics, remote sensing and field observations for assessing soil loss in challenging drylands: A case study of Northwestern Somalia. Land Degrad. Dev. 20, 101–115.
Parplies, A., Dubovyk, O., Tewes, A., Mund, J.-P., Schellberg, J., 2016. Phenomapping of rangelands in South Africa using time series of RapidEye data. Int. J. Appl. Earth Obs. Geoinf. 53, 90–102.
Peerbhay, K., Mutanga, O., Lottering, R., Bangamwabo, V., Ismail, R., 2016. Detecting bugweed *(Solanum mauritianum)* abundance in plantation forestry using multisource remote sensing. ISPRS J. Photogramm. Remote Sens. 121, 167–176.
Perroy, R.L., Sullivan, T., Stephenson, N., 2017. Assessing the impacts of canopy openness and flight parameters on detecting a sub-canopy tropical invasive plant using a small unmanned aerial system. ISPRS J. Photogramm. Remote Sens. 125, 174–183.
Peterson, E.B., 2005. Estimating cover of an invasive grass (Bromus tectorum) using tobit regression and phenology derived from two dates of Landsat ETM+ data. Int. J. Remote Sens. 26, 2491–2507.
Platts, P.J., Omeny, P.A., Marchant, R., 2015. AFRICLIM: high-resolution climate projections for ecological applications in Africa. Afr. J. Ecol. 53, 103–108.
Powers, D.M.W., 2011. Evaluation: from precision, recall and F-measure to ROC informedness, markedness & correlation. J. Mach. Learn. Tech. 2, 37–63.
Reunanen, J., 2003. Overfitting in making comparisons between variables selection methods. J. Mach. Learn. Res. 3, 1371–1382.
Rodriguez-Galiano, V.F., Ghimire, B., Rogan, J., Chica-Olmo, M., Rigol-Sanchez, J.P., 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS J. Photogramm. Remote Sens. 67, 93–104.
Runnström, M.C., 2000. Is Northern China Winning the Battle against Desertification? AMBIO. J. Hum. Environ. 29, 468–476.
Schwarz, M., Zimmermann, N.E., 2005. A new GLM-based method for mapping tree cover continuous fields using regional MODIS reflectance data. Remote Sens. Environ. 95, 428–443.
Senf, C., van der Pflugmacher, D., Linden, S., Hostert, S., 2013. Mapping rubber plantations and natural forests in Xishuangbanna (Southwest China) using multi-Spectral phenological metrics from MODIS time series. Remote Sens. 5, 2795–2812.
Seta, T., Assefa, A., Mesfin, F., Balcha, A., 2013. Distribution status and the impact of parthenium weed (*Parthenium hysterophorus* L.) at Gedeo Zone (Southern Ethiopia). Afr. J. Agric. Res. 8, 386–397.
Shiferaw, H., Teketay, D., Nemomissa, S., Assefa, F., 2004. Some biological characteristics that foster the invasion of *Prosopis juliflora* (Sw.) DC. at Middle Awash Rift Valley Area, north-eastern Ethiopia. J. Arid Environ. 58, 135–154.
Shitanda, D., Mukonyi, K., Kagiri, M., Gichua, M., Simiyu, L., 2014. Properties of Prosopis juliflora and its potential uses in ASAL areas of Kenya. J. Agric. Sci. Technol. 15, 15–27.
Singh, K.K., Chen, Y.-H., Smart, L., Gray, J., Meentemeyer, R.K., 2018. Intra-annual phenology for detecting understory plant invasion in urban forests. ISPRS J. Photogramm. Remote Sens. 142, 151–161.
Singh, K.K., Davis, A.J., Meentemeyer, R.K., 2015. Detecting understory plant invasion in urban forests using LiDAR. Int. J. Appl. Earth Obs. Geoinf. 38, 267–279.
Tabachnick, B.G., Fidell, L.S., Osterlind, S.J., 2001. Using Multivariate Statistics, 4th ed. Allyn & Bacon, Inc., Massachusetts.
Tabu, J.S., Otwelo, J.A., Koskei, P., Makokha, P., 2013. Hazard analysis of arid and semi-arid (ASAL) regions of Kenya. East Afr. J. Public Health. 10, 411–416.
Tamado, T., Milberg, P., 2000. Weed flora in arable fields of eastern Ethiopia with emphasis on the occurrence of Par*thenium hysterophorus*. Weed Res. 40, 507–521.
Tarantino, C., Casella, F., Adamo, M., Lucas, R., Beierkuhnlein, C., Blonda, P., 2019. *Ailanthus altissima* mapping from multi-temporal very high resolution satellite images. ISPRS J. Photogramm. Remote Sens. 147, 90–103.
Taylor, S.L., Hill, R.A., Edwards, C., 2013. Characterising invasive non-native Rhododendron ponticum spectra signatures with spectroradiometry in the laboratory and field: Potential for remote mapping. ISPRS J. Photogramm. Remote Sens. 81, 70–81.
Vargas, R.R., Omuto, C.T., Alim, M.S., Ismail, A., Njeru, L., 2009. Land degradation assessment and recommendation for a monitoring framework in Somaliland. Somalia Water and Land Information Management, Nairobi.
Vilà, M., Espinar, J.L., Hejda, M., Hulme, P.E., Jarošík, V., Maron, J.L., Pergl, J., Schaffner, U., Sun, Y., Pyšek, P., 2011. Ecological impacts of invasive alien plants: a meta-analysis of their effects on species, communities and ecosystems. Ecol. Lett. 14, 702–708.
Vrieling, A., de Leeuw, J., Said, M.Y., 2013. Length of growing period over Africa: Variability and trends from 30 years of NDVI time series. Remote Sens. 5, 982–1000.
Wallace, C.S.A., Walker, J.J., Skirvin, S.M., Patrick-Birdwell, C., Jake, F., Weltzin, J.F., Raichle, H., 2016a. Mapping presence and predicting phenological status of invasive buffelgrass in southern Arizona using MODIS, climate and citizen science observation data. Remote Sens. 8, 524.
Wallace, C.S.A., Walker, J.J., Skirvin, S.M., Patrick-Birdwell, C., Weltzin, J.F., Raichle, H., 2016b. Mapping presence and predicting phenological status of invasive buffelgrass in southern Arizona using MODIS, climate and citizen science observation data. Remote Sens. 8, 524.
Wasserstein, R.L., Lazar, N.A., 2016. The ASA's statement on *p*-values: context, process, and purpose. Am Stat. 70, 129–133.
Wasserstein, R.L., Schirm, A.L., Lazar, N.A., 2019. Moving to a world beyond "$p < 0.05$". Am Stat. 73, 1–19.
Wessels, K.J., van den Bergh, F., Scholes, R.J., 2012. Limits to detectability of land degradation by trend analysis of vegetation index data. Remote Sens. Environ. 125, 10–22.
West, A.M., Evangelistaa, P.H., Jarnevichd, C.S., Kumar, S., Swallowe, A., Luizzaa, M.W., Chignell, S.M., 2017. Using multi-date satellite imagery to monitor invasive grass species distribution in post-wildfire landscapes: An iterative, adaptable approach that employs open-source data and software. Int. J. Appl. Earth Obs. Geoinf. 59, 135–146.
Xie, Z., Roberts, C., Johnson, B., 2008. Object-based target search using remotely sensed data: A case study in detecting invasive exotic Australian Pine in south Florida. ISPRS J. Photogramm. Remote Sens. 63, 647–660.
Zhang, X., Friedl, M.A., Schaaf, C.B., 2006. Global vegetation phenology from Moderate Resolution Imaging Spectroradiometer (MODIS): Evaluation of global patterns and comparison with in situ measurements. J. Geophys. Res. 111, G04017.
Zuberi, M.I., Gosaye, T., Hossain, S., 2014. Potential threat of allien invansive species: *Parthenium hysterophorus* L. to subsistance agriculture in Ethiopia. Sarhad J. Agric. 30, 117–125.