

RESEARCH

Open Access



Species Distribution Modelling performance and its implication for Sentinel-2-based prediction of invasive *Prosopis juliflora* in lower Awash River basin, Ethiopia

Nurhussen Ahmed^{1,2*} , Clement Atzberger³ and Worku Zewdie¹

Abstract

Background: Species Distribution Modelling (SDM) coupled with freely available multispectral imagery from Sentinel-2 (S2) satellite provides an immense contribution in monitoring invasive species. However, attempts to evaluate the performances of SDMs using S2 spectral bands and S2 Radiometric Indices (S2-RIs) and biophysical variables, in particular, were limited. Hence, this study aimed at evaluating the performance of six commonly used SDMs and one ensemble model for S2-based variables in modelling the current distribution of *Prosopis juliflora* in the lower Awash River basin, Ethiopia. Thirty-five variables were computed from Sentinel-2B level-2A, and out of the variables, twelve significant variables were selected using Variable Inflation Factor (VIF). A total of 680 presence and absence data were collected to train and validate variables using the tenfold bootstrap replication approach in the R software “sdm” package. The performance of the models was evaluated using sensitivity, specificity, True Skill Statistics (TSS), kappa coefficient, area under the curve (AUC), and correlation.

Results: Our findings demonstrated that except bioclim all machine learning and regression models provided successful prediction. Among the tested models, Random Forest (RF) performed better with 93% TSS and 99% AUC followed by Boosted Regression Trees (BRT), ensemble, Generalized Additive Model (GAM), Support Vector Machine (SVM), and Generalized Linear Model (GLM) in decreasing order. The relative influence of vegetation indices was the highest followed by soil indices, biophysical variables, and water indices in decreasing order. According to RF prediction, 16.14% (1553.5 km²) of the study area was invaded by the alien species.

Conclusions: Our results highlighted that S2-RIs and biophysical variables combined with machine learning and regression models have a higher capacity to model invasive species distribution. Besides, the use of machine learning algorithms such as RF algorithm is highly essential for remote sensing-based invasive SDM.

Keywords: Biophysical processors, *Prosopis juliflora*, Radiometric indices, Remote sensing, Sentinel-2, Species Distribution Modelling

* Correspondence: nurhussenahmed40@gmail.com

¹Department of Remote Sensing, Entoto Observatory and Research Center, Ethiopian Space Science and Technology Institute (ESSTI), Addis Ababa, Ethiopia

²Department of Geography and Environmental Studies, Wollo University, Dessie, Ethiopia

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Introduction

Globally, the invasiveness of invasive alien plant species has become a great risk as it adversely affects ecological services and socio-economic systems (Rajah et al. 2019; Paz-Kagan et al. 2019). The distribution and subsequent socio-economic loss in East Africa are also increasing at an alarming rate (Landmann et al. 2020). Despite their current distribution and negative consequences, studies also daunt an expected large increase in their invasion and will adversely affect uninvaded areas (Howard 2019). Hence, monitoring the adverse impacts of invasive alien plant species using contemporary technologies before its dispersion has paramount importance (Rajah et al. 2019; West et al. 2016). This is particularly vital for developing countries with no or little financial and technical capabilities to avert the invasion and any delay further aggravates the problem (Pyšek et al. 2012; Vilà et al. 2011).

Prosopis juliflora (hereafter *Prosopis*) is one of the ten worst invasive species that adversely affect millions of hectares of land in many arid and semi-arid regions (Ilukor et al. 2016; Rembold et al. 2015; Shackleton et al. 2014). In Ethiopia, about 1.17 million hectares of land are currently invaded by *Prosopis* only in the Afar region, which results in approximately 602 million US dollar loss of ecosystem service (mainly due to *Prosopis* expansion) (Shiferaw et al. 2019b). Its social, economic, and ecological adverse impact in the area will also be expected to increase as the species is aggressively increasing at a rate of 8.3% annually (Shiferaw et al. 2019b). Hence, the best and recommended way is to control its invasion early by anticipating suitable habitats for its diversification using timely and cost-effective tools (Reaser et al. 2020; West et al. 2017).

Mapping the current invasion and modelling suitable habitat of invasive species has an immense contribution for ecologists and policymakers to control the expansion and its adverse threat (Ayanu et al. 2014; Evangelista et al. 2008; Feilhauer et al. 2012). However, controlling its spread and managing its consequence needs a robust, cost-effective, efficient, and precise monitoring system (Lopatin et al. 2016). In addition, ecologists are also in need of timely and cost-effective methods to model and predict in advance the distribution of invasive species (West et al. 2017). In this regard, the use of Species Distribution Modelling (SDM) and Geographic Information System (GIS) are among the widely used prediction tools (Bradley 2014). SDM's had been used by ecologists for a long time to predict species distribution (Allouche et al. 2006; Jiménez-Valverde 2014; Lemke and Brown 2012; Wisz et al. 2008). However, models have shown varied performance and no single best model has been identified by studies for different species and environments (Reside et al. 2011).

The selection and implementation of models require great care as inappropriate use of models can affect the accuracy of species prediction (Elith and Leathwick 2009; Elith et al. 2006). Consequently, many studies used more than one model in comparison (Früh et al. 2018; Ng et al. 2018; Stohlgren et al. 2010). Besides, some studies recommend the use of presence-absence models, while others appreciate the use of presence-only models (Elith and Leathwick 2009). Owing to this, many studies compared machine learning algorithms with regression-based models (Shiferaw et al. 2019c) while others made a comparison among machine learning algorithms (Früh et al. 2018) and others also developed ensemble models from different SDMs rather than relying on a single model (Ng et al. 2018). However, the performances of SDMs are dependent upon the type of application and problem (Bhattacharya 2013), the spatial resolution of environmental variables (Reside et al. 2011), and the selection of environmental predictors (Elith and Leathwick 2009). Owing to this, the use of several models by applying similar environmental variables and methodologies gives confidence for ecologists to judge their results (Fischer et al. 2013; Lemke and Brown 2012).

The use of machine learning algorithms for remotely sensed-based prediction gets recent attention and significantly improves the prediction of invasive species (Benito et al. 2013; Früh et al. 2018). A timely, advanced, and cost-effective approach that integrates remote sensing technology to monitor invasive species risk is highly needed (Rajah et al. 2019; West et al. 2017). This is particularly important in arid and semi-arid regions of developing nations as the cost of survey data and high-resolution commercial data is difficult to justify. The possible options in such areas are the use of freely available multispectral data (Jensen et al. 2020). In this regard, S2 data due to its high spatial, spectral, and temporal resolutions provide an immense contribution to monitoring the distribution and spread of invasive species (Martinez et al. 2020; Meroni et al. 2017; Ng et al. 2016, 2017; Truong et al. 2017). S2's high spectral resolution allowed ecologists to derive numerous indices (Rajah et al. 2019). The use of these indices is better than raw bands as they can reduce the effect of atmospheric condition and soil background on canopy reflectance (Liu et al. 2005). Among the available indices, most studies usually employed S2-derived Vegetation Indices (S2VIs). Though S2VIs have higher importance, the contribution of other radiometric indices is potentially also very high. For example, soil radiometric indices (Nouri et al. 2018) and biophysical variables have a higher capacity in monitoring and managing vegetation changes (Atzberger et al. 2015; Mudereri et al. 2019). Hence, evaluating the contribution of S2-derived RIs is of utmost importance as they incorporate different variables

from the soil, water vegetation, and biophysical characteristics of the area.

So far, a number of studies have evaluated the performance of SDMs for mapping and modelling different invasive species. Stohlgren et al. (2010) made a comparison of five individual models for the prediction of invasive species and suggested that the use of an ensemble model significantly improves prediction compared to a single model. Früh et al. (2018) made a comparison on performances of four machine learning models in the prediction of invasive species and recommended the use of an ensemble approach from best-performing models rather than a single or ensemble model from all models. Likewise, Ng et al. (2018) made a comparison among machine learning models and argued that the performance of Random Forest (RF) and ensemble models are highly appreciated compared to other models used in their study. Abdi (2020) made a comparison of four machine learning algorithms using S2-derived variables for land cover classification and concluded that the Support Vector Machine (SVM) performed better than other models. These studies have used either integration of remote sensing and non-remote sensing datasets (e.g., Ng et al. 2018) or used coarse resolution remote sensing data (Stohlgren et al. 2010). Other studies used S2 data for land cover classification (Abdi 2020), mapping, and detecting invasive species distribution (Rajah et al. 2019). In Ethiopia, few studies were conducted on mapping and prediction of *Prosopis*. Wakie et al. (2014) assess the distribution of invasive *Prosopis* using Moderate Resolution Imaging Spectroradiometer (MODIS) data and Maxent model. Besides, Shiferaw et al. (2019c) evaluated the performance of different SDM using Landsat 8 Operational Land Manager (OLI), climate, and infrastructural data. However, comparative studies about SDM using high-resolution data in general and S2-RIs and biophysical variables in particular for monitoring *Prosopis* were scarce.

This study, therefore, aims at addressing the following research gaps and needs: (1) identifying a robust method for modelling remote sensing-based invasive species distribution is highly required; (2) assessing the potential of S2-derived vegetation, soil and water indices, and biophysical variables for modelling invasive species distribution in arid and semi-arid regions of developing countries is highly essential.

Materials and methods

Study area and species

The study was carried out in the lower Awash River basin, Ethiopia. It is located between 40.74 to 41.82° longitude and 10.99 to 12.36° latitude (Fig. 1). It covers an area of 9471.5 km² with elevations ranging from 240 to 1341 m above mean sea level. In addition, 75% (7103.6

km²) of the study area is found in the desert and 25% (2367.9 km²) is found in arid to semi-arid agro-ecological zones. It is also part of the great east African rift valley system. According to the National Meteorological Agency (NMA) (2020), the mean annual temperature, mean maximum, and mean minimum annual temperature at Dubti station for years between 2002 and 2017 were recorded as 28.1 °C, 33.5 °C, and 22.6 °C, respectively. December and July are the coldest and warmest months, respectively. Furthermore, frequent drought and subsequent famine are the two major characteristics of the area (Mulugeta et al. 2019).

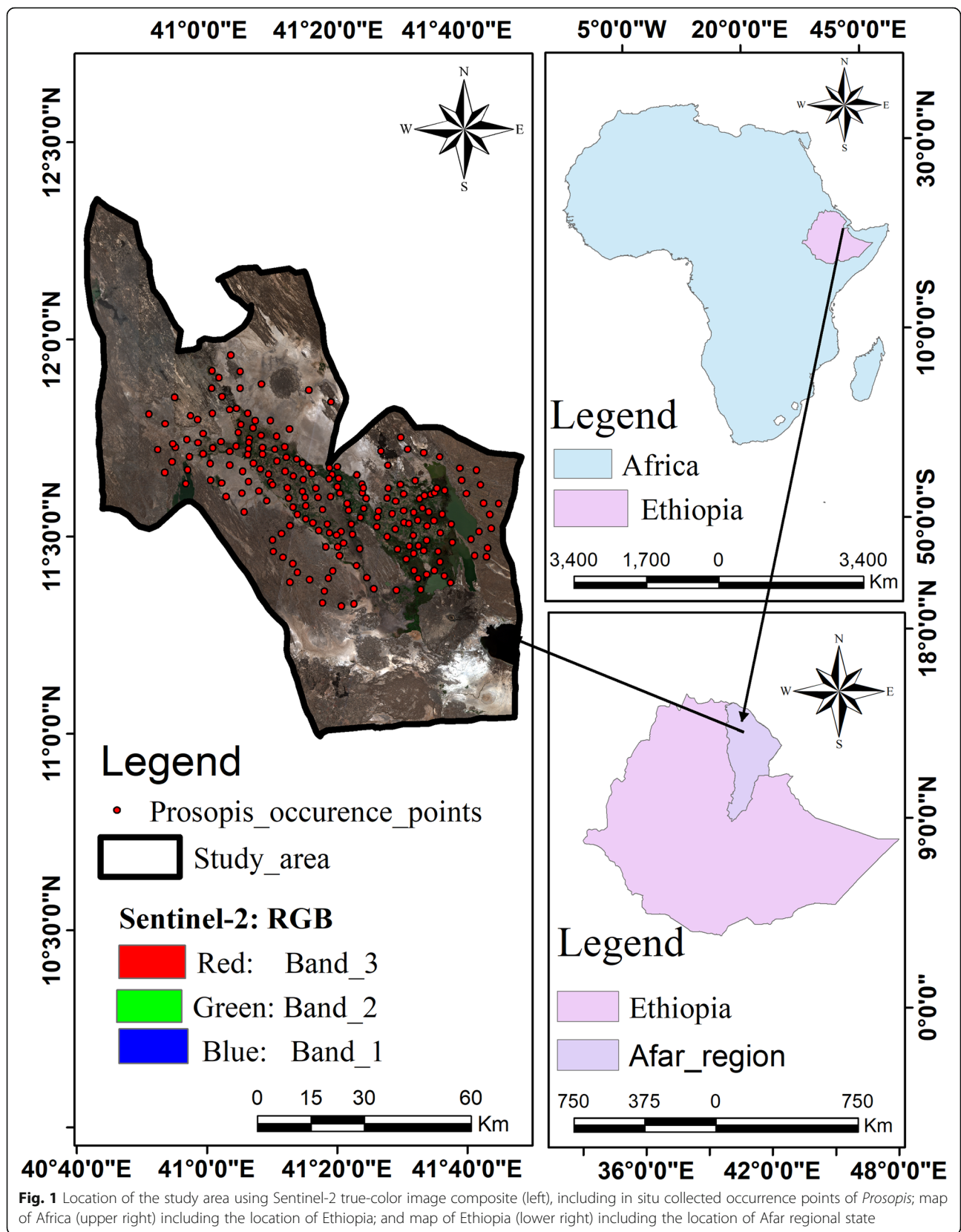
Though pastoralism is the dominant way of life, agro-pastoralism is also practicing in the area. The state-owned Tendaho irrigation project which covers around 62,500 ha along the lower Awash River basin supports irrigation-based agriculture (Tadese et al. 2019). The cultivation of sugarcane, wheat, cotton, maize, and other vegetables has been also practiced in small-scale agriculture. In the Afar region, *Prosopis* was introduced in the early 1980s for soil and water conservation (Tilahun et al. 2017). This was part of the then government afforestation initiative to combat drought and desertification (Wakie et al. 2014). Before the invasion of *Prosopis*, native grasses, forbs, shrubs, and woody plants dominantly covered the area and were an important source of fodder for the locality (Ayanu et al. 2014; Wakie et al. 2014). After the invasion of the species, however, conflict among pastoralists has increased due to resource competition (Ilukor et al. 2016; Mehari 2015; Wakie et al. 2012). The expansion of *Prosopis* also affected livestock production with an expected loss of about 26 million dollars per year in the region (Ilukor et al. 2014).

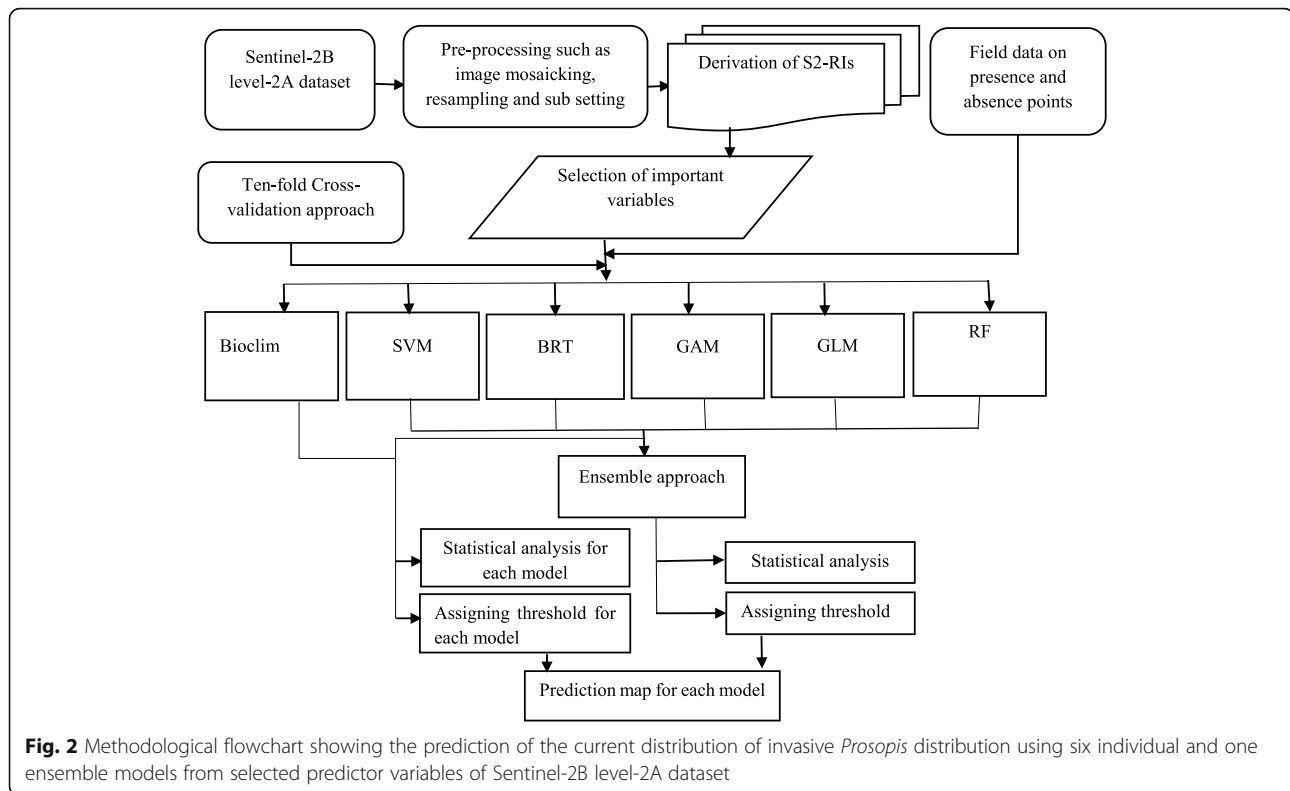
Method overview

In this study, we evaluated the performance of six commonly used SDMs using S2-RIs. We computed thirty-five variables from vegetation, soil, water radiometric indices, and biophysical variables. Out of the variables, twelve significant variables were selected using the VIF. TSS, AUC, kappa, correlation, sensitivity, and specificity were used to evaluate model performance. Besides, reference data on the presence and absence of the species were collected using a handheld Global Positioning System (GPS). Also, an ensemble model was developed from best-performed models (except the least performed bioclim model). At last, prediction maps for all individual and ensemble models were produced. Graphical illustrations on the overall methods used for this study are presented in Fig. 2.

Presence and absence data

Georeferenced in situ data were collected with the help of a handheld GPS from both presence (invaded) and absence (uninvaded) points in the dry season between





January and February 2020. In this period, *Prosopis* is highly discriminated from the other tree species as most tree types shed their leaves due to water scarcity (Godoy et al. 2011; Meroni et al. 2017; Xu 2014). A total of 680 points were collected using stratified random sampling in a 10 m by 10 m plot similar to the spatial resolution of S2 data (Arogroundade et al. 2020). Of all reference points, 30% (205 points) were presence while 70% (475 points) were absence data. This share was chosen considering the previous distribution of *Prosopis* in the area (Linders et al. 2020). Field data were collected throughout the study area and 200 m was the minimum distance among the points.

To evaluate the spatial autocorrelation among observations, Moran's I was used. Accordingly, we found 0.28 of Moran's Index and 2.45 of Z -score, indicating no apparent spatial clustering among the points (Abdulhafedh 2017). To further reduce spatial autocorrelation among points, we used the "Spatially Rarefy Occurrence Data for SDMs (reduce spatial autocorrelation)" function in ArcGIS SDM Toolbox. To get independent validation statistics, 70% of the collected data were used to train models while 30% were used to validate models (Engler et al. 2013).

Satellite image processing and selection of variables

In this study, Sentinel-2B level-2A data was used to produce varied radiometric indices and biophysical

variables. Sentinel-2B level-2A product provides geometric and radiometrically corrected images. The data is delivered as Bottom-Of-Atmosphere (BOA) reflectance images converted from the level-1C Top-Of-Atmosphere (TOA) reflectance (Szantoi and Strobl 2019). Its application was tested by Vuolo et al. (2016) and used by different studies (e.g., Arogroundade et al. 2020; Ng et al. 2018) in the areas of invasive species distribution. It can detect the Earth's surface at 10-m, 20-m, and 60-m spatial resolutions. Sentinel-2B level-2A product acquired between 19 January 2020 and 28 February 2020 which is concurrent with the field data collection campaign were downloaded from the European Space Agency (ESA) data portal (<https://scihub.copernicus.eu/dhus/#/home>).

A total of four scenes were required to cover the study area. Pre-processing such as image mosaicking, resampling to a common grid of 10-m resolution, and sub-setting were made using freely available Sentinel Application (SNAP) 7.0 software. Maps were produced by using ArcGIS software. A total of thirty-five variables (Table S1): seventeen from vegetation radiometric indices, eight from soil radiometric indices, five from water radiometric indices, and five from biophysical variables were considered (Table S1). To select important variables, we used the VIF correlation approach to reduce multicollinearity problems in the R 4.0 software "sdm" package (Naimi and Araújo 2016). VIF was used by

Table 1 Description of the predictive variables used in modelling and mapping the current distribution of invasive *Prosopis*. The variables were categorized under vegetation indices, soil indices, water radiometric indices, and biophysical variables

Variables	Short description
Brightness Index (BI)	BI is a soil radiometric index that represents the mean of the brightness of satellite images. It is highly associated with soil brightness (Escadafal 1989).
Colour Index (CI)	CI is a soil radiometric index that helps to differentiate soils and their development. Higher CI values representing crusted soils and sands while lower CI values indicating a high concentration of carbonates or sulfates (Escadafal 1989).
Canopy Water Content (CWC)	CWC is a biophysical variable that quantifies the amount of water in the given area. It is also an essential predictor in the areas of agriculture and forestry (Cernicharo et al. 2013).
Fraction of Absorbed Photosynthetically Active Radiation (FAPAR)	FAPAR is an important biophysical variable that indicates the capacity of the vegetation canopy to absorb Photosynthetically Active Radiation (Fensholt et al. 2004).
Leaf Area Index (LAI)	LAI is a key biophysical variable that quantifies the amount of leaf area per unit ground surface (Zhang et al. 2003).
Modified Chlorophyll Absorption Ratio Index (MCARI)	MCARI is a vegetation radiometric index, responsive to chlorophyll variations (Daughtry et al. 2000).
Modified Normalized Difference Water Index (MNDWI)	MNDWI is a water radiometric index promoted to enhance open water features by minimizing the effect of vegetation, soil, and built-up land noises (Xu 2006).
Meris Terrestrial Chlorophyll Index (MTCI)	MTCI is a water radiometric index that estimates the amount of chlorophyll (Dash and Curran 2004).
Redness Index (RI)	RI is a soil radiometric index that gives information about soil color variation in a given area. It is an important index to measure soil redness in the arid environment (Mathieu et al. 1998).
Sentinel-2 Red-Edge Position Index (S2REP)	S2REP is a vegetation radiometric index that provides information on chlorophyll content and the growth status of vegetation.
Transformed Normalized Difference Vegetation Index (TNDVI)	TNDVI is a vegetation radiometric index that shows the amount of green biomass in a pixel. It has a high coefficient of determination and excellent linearity to vegetation cover (Bannari et al. 2002).
Transformed Soil Adjusted Vegetation Index (TSAVI)	TSAVI is a soil radiometric index developed to minimize the influence of soil brightness (Baret and Guyot 1991).

several studies in the areas of invasive species distribution as a tool to select important variables (Ng et al. 2018; Zimmermann et al. 2007). Accordingly, for this study, out of thirty-five variables, a total of twelve important variables (Table 1) with threshold values less than 0.7 were selected (Engler et al. 2013; Zimmermann et al. 2007). Furthermore, the relative importance of variables for all models was computed using the “getVarImp” function in the R software “SDM” package (Naimi

and Araújo 2016). Then, the weighted mean values of each variable for each model run was calculated and categorized as vegetation, soil radiometric indices, and biophysical variables.

Selection of Species Distribution Modelling

Today a large number of modelling methods are available and can be classified as “profile,” “regression,” and “machine learning” (Hijmans and Elith 2019). This study

Table 2 Predictive models from machine learning (RF, SVM, and BRT), regression (GAM and GLM), and profile (bioclim) methods and their short description and common pieces of literature that have used these models in modelling invasive species distribution

Models	Short description	Examples
Random Forest (RF)	RF (Breiman 2001), a combination of tree predictors, is the most commonly used machine learning algorithm (Abdi 2020). It is an effective method for predicting species richness and density (Kosicki 2020).	Früh et al. (2018); Mi et al. (2017); Ng et al. (2018); West et al. (2017)
Support Vector Machine (SVM)	SVM (Cortes and Vapnik 1995) is characterized by its ability to generalize features. It can be used for classification and regression-based studies.	Abdi (2020); Früh et al. (2018); Ng et al. (2016)
Boosted Regression Trees (BRT)	Like RF, BRT is based on a combination of a relatively small number of trees to increase the performance of predictive variables (Elith et al. 2008). It has also the capacity to process several predictors at high predictive accuracy (Gu et al. 2019).	Guisan et al. (2007); West et al. (2017)
Generalized Additive Model (GAM)	GAM is a popular regression model and is extensively used in ecological studies for modelling non-normal distributions (Ravindra et al. 2019).	Guisan et al. (2007); Soultan and Safi (2017)
Generalized Linear Model (GLM)	GLM can process and manage non-linear data structures. Its flexibility makes it better suited for ecological-based studies (Guisan et al. 2002).	Guisan et al. (2007); Soultan and Safi (2017); West et al. (2017)
Bioclim	Bioclim was the earliest SDM package used in many ecological studies including invasive species prediction (Booth et al. 2014).	Guisan et al. (2007); Hernandez et al. (2006); Reiss et al. (2011)

evaluates the performance of six commonly used models in the areas of invasive SDM. The models were selected from machine learning algorithms, regression models, and profile methods for comparison reasons (Table 2). Boosted Regression Trees (BRT), RF, and SVM from machine learning models; Generalized Additive Model (GAM) and Generalized Linear Model (GLM) from regression models; and bioclim from profile methods were considered. One ensemble model from all models, except the least performing bioclim model, was also developed.

Model validation and mapping

For validations of the above-listed models, we used the bootstrap replication approach in the R 4.0 software “SDM” package developed by Naimi and Araújo (2016). Out of 680 collected points, 205 (30%) randomly selected points were used as test data to validate models and the remaining 70% were used for the training of models. This step was replicated ten times and their mean values of sensitivity, specificity, TSS, kappa, AUC, and correlation were used to assess the accuracy of the models. The bootstrapping replication method has the potential to offer unbiased predictive accuracy with fairly low variance (Harrell et al. 1996; Lima et al. 2019). Besides, the sensitivity-specificity sum maximization approach was used to select the best threshold. This threshold was recommended as the best approach for the prediction of species distribution (Liu et al. 2005).

Binary maps were developed as pixels greater than the threshold represented the presence of *Prosopis* at different levels of invasion and pixels lower than the threshold indicated the absence of *Prosopis* in the area for all models. Besides, a correction for over-prediction using clip models by buffered Minimum Convex Polygon (MCP) was made in ArcGIS SDM Toolbox. The buffered MCP as a posteriori method enables the reduction of over-prediction (Mendes et al. 2020). In addition, the ensemble model was evaluated using a weighted mean of all models except the least performing bioclim model. Maps were further classified into “uninvaded,” “low-

invasion,” “moderate-invasion,” and “high-invasion” for the ensemble model.

Results

Performances of Species Distribution Modelling

The performances of SDMs using different evaluation techniques are presented in Table 3. Except for the bioclim model, the accuracy of the other models was very high. Machine learning algorithms (RF and BRT) performed better than regression models (GLM and GAM) and profile (bioclim) method for all evaluation techniques. Following RF, BRT, Ensemble, GAM, SVM, GLM, and bioclim performed better in decreasing order. In addition, BRT performed well in AUC and sensitivity, and GAM performed well in kappa and correlation evaluation techniques after RF.

Model accuracy can also be evaluated using the receiver operator characteristics (ROC) curve as it has the capacity to show the proportion of the true presence rate (sensitivity) and the true absence rate (specificity). The ROC curve for all models is presented in Fig. 3. Except in bioclim, sensitivity and specificity scores were very high for all models indicating both invaded and uninvaded areas were well identified and the proportion of correctly classified samples were maximum (Fig. 3).

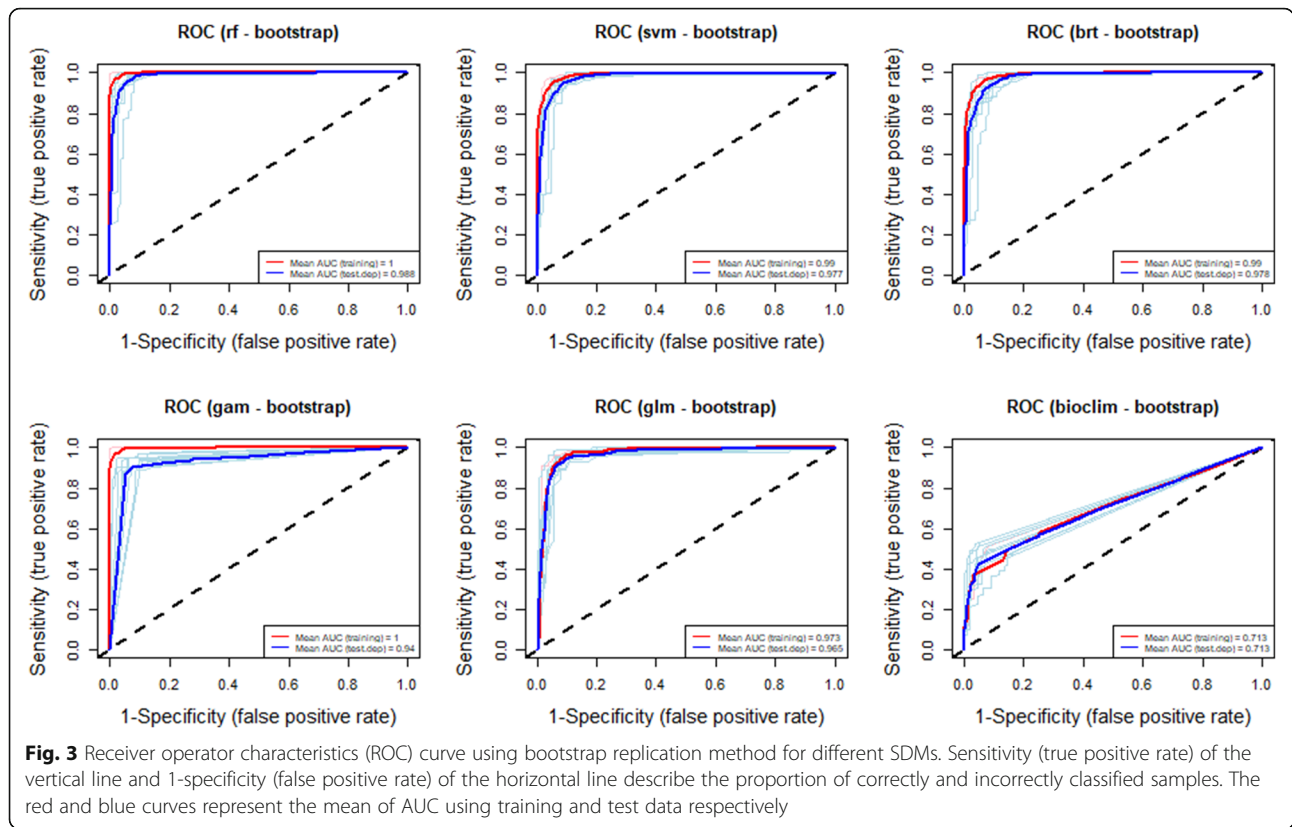
Prosopis distribution

The presence (invaded) of *Prosopis* distribution for RF, SVM, GLM, BRT, GAM, Ensemble, and bioclim models were 14.3%, 12.9%, 12.8%, 13.9%, 14.9%, 16.11%, and 3.9%, respectively (Fig. 4). According to the best-performed RF model, 1354.6 km² of the study area was invaded by *Prosopis*, which was more intense around the central part (Fig. 4).

Furthermore, the ensemble model was used to produce maps showing the invasion of *Prosopis* at varying levels of distribution. The threshold for the ensemble model was 0.47 (Table 3), and pixels below the threshold were considered as “uninvaded” of *Prosopis* and pixel values above the threshold were further divided into three classes as “low,” “medium,” and “high” invasion of

Table 3 Performance evaluation of SDMs using different statistical parameters. Sensitivity and specificity describe the rate of true positive and negative respectively

Models	TSS	Kappa	Correlation	AUC	Sensitivity	Specificity	Threshold
RF	0.93	0.91	0.93	0.99	0.98	0.95	0.46
SVM	0.89	0.86	0.89	0.98	0.95	0.91	0.52
BRT	0.90	0.88	0.90	0.98	0.97	0.90	0.49
GLM	0.87	0.86	0.87	0.97	0.94	0.92	0.47
GAM	0.90	0.90	0.89	0.96	0.96	0.93	0.41
Bioclim	0.37	0.34	0.41	0.69	0.42	0.94	0.40
Ensemble	0.90	0.88	0.90	0.98	0.96	0.92	0.47



Prosopis distribution (Ng et al. 2018). Accordingly, 86.8% of the study place was not invaded by *Prosopis* distribution. The rest of the study place (13.2%) was invaded by *Prosopis* at different levels of invasion as low (2.5%), medium (3.2%), and high (7.5%) distribution (Fig. 5).

Relative contribution of predictor variables

The relative influence of predictors is shown in Fig. 6 and supporting information (Table S2). The relative influence of few variables was very high while other variables were found to be insignificant. The relative influence of vegetation radiometric indices (TNDVI, MCARI, MTCL, and S2REP) for RF, SVM, BRT, GLM, GAM, and bioclim were 83%, 65.75%, 74.35%, 75.5%, 54.85%, and 51.95%, respectively. However, the relative importance of water radiometric indices (MNDWI) was the least in all models except in the bioclim model (Fig. 6 and Table S2).

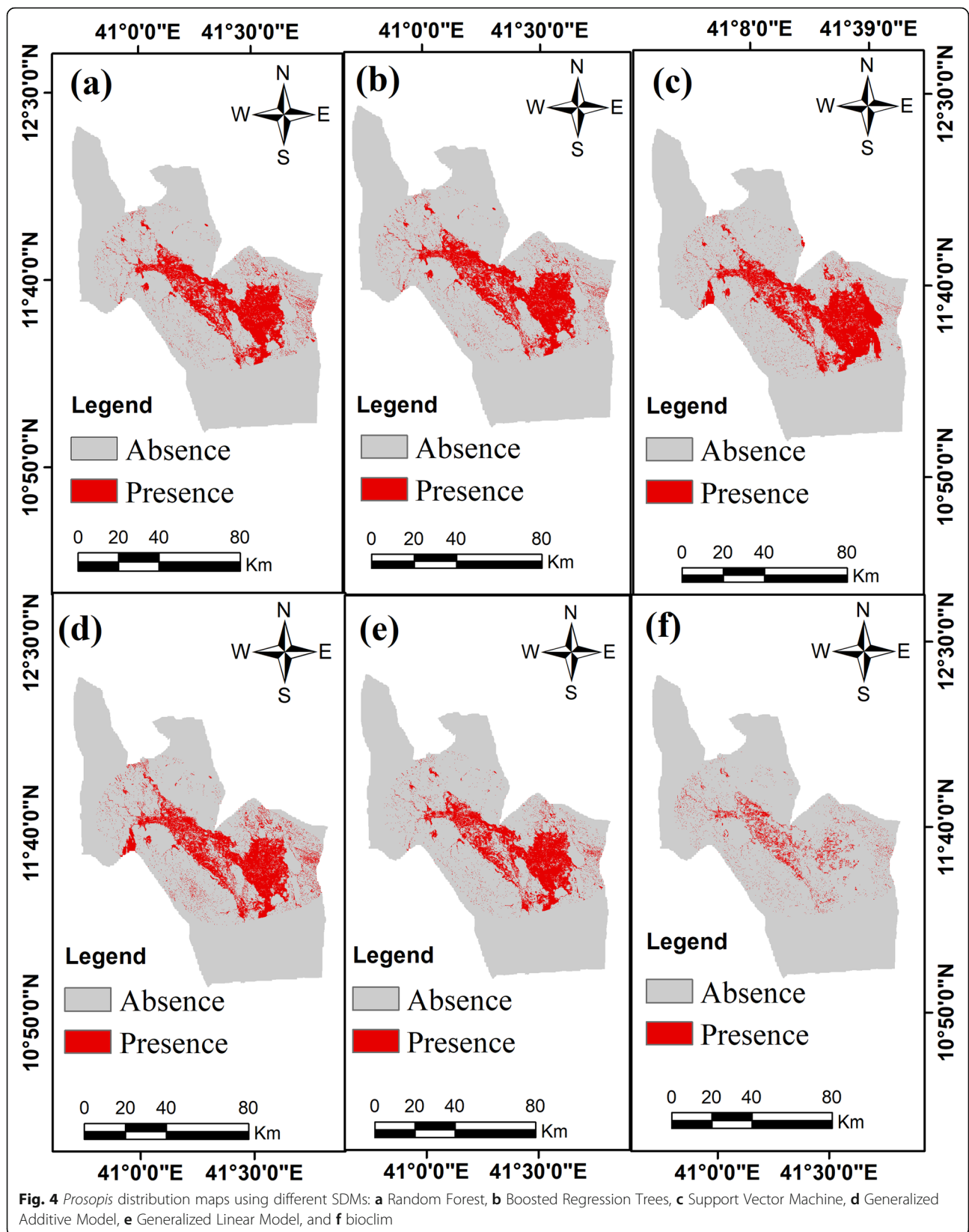
Discussion

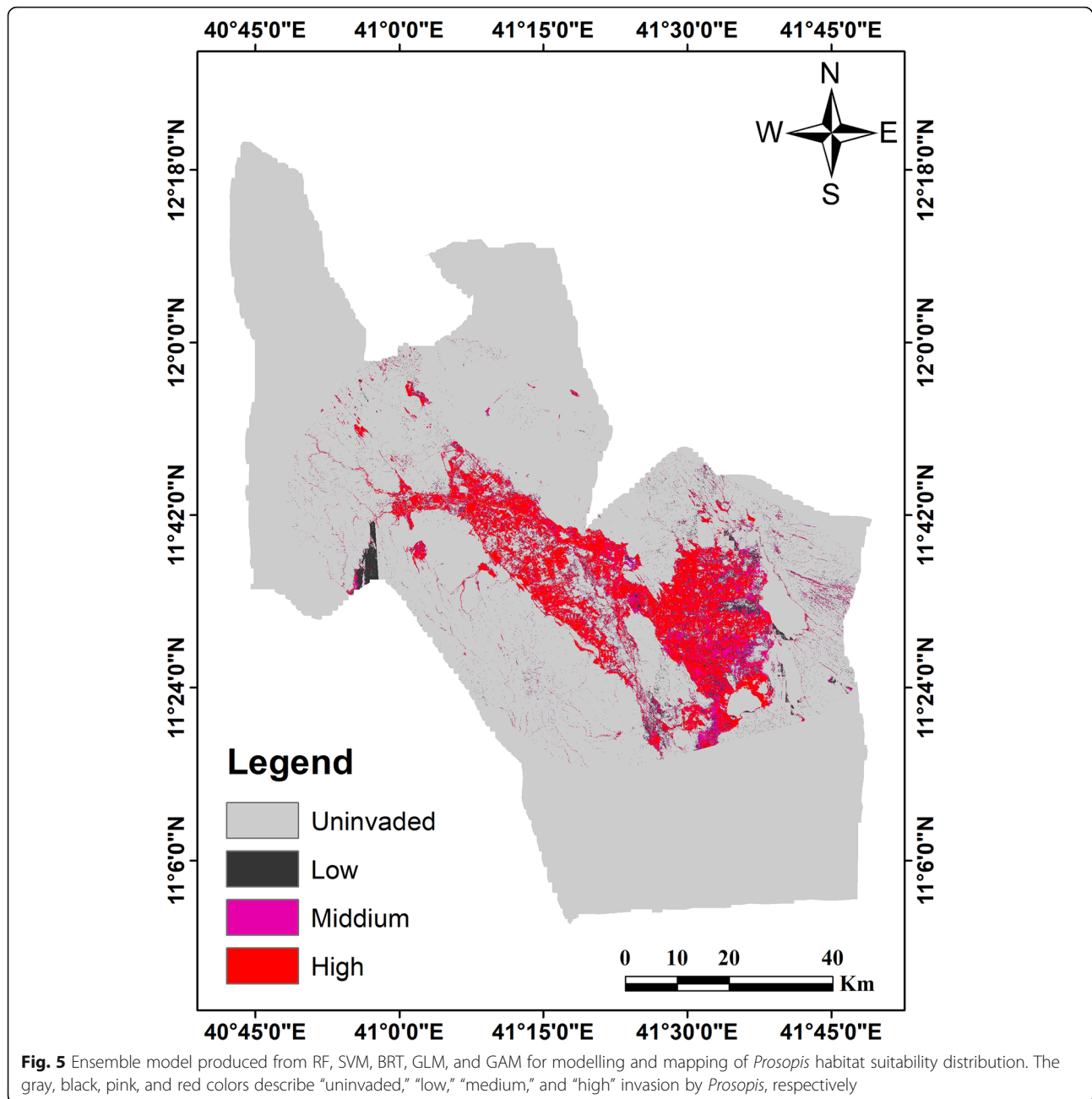
Implications of SDMs performance for remote sensing-based prediction

Our study highlights the relative performance of SDMs for remote sensing-based prediction of invasive species distribution. In the present study, the higher performance of machine learning algorithms (RF, BRT, and SVM) and regression models (GLM and GAM) were

observed. Among all models, the bioclim model performed worst. The result of our study varied from 19.75% (GAM) to 5% (bioclim) prediction. This huge difference between the models' predictions can affect the monitoring of invasive species. According to González-Ferreras et al. (2016), models with AUC values below 0.7 and above 0.9 are considered as "very poor" and "highly accurate," respectively. Besides, models with TSS and kappa values below 0.4 and above 0.8 are considered as "poor" and "excellent," respectively. Based on the above evaluation techniques the performances of all models, except bioclim, were in the category of "excellent" whereas the performance of the bioclim model was in the category of "very poor" and "poor." The prediction obtained from the above models, except for bioclim, also agreed with previous studies conducted in the area indicating the performance of SDMs has great implication in providing certain predictions. Moreover, numerous studies highlight the higher performance of RF for remote sensing-based prediction of invasive species distribution after evaluating the performance of several SDMs (Jensen et al. 2020; West et al. 2016, 2017).

A study by Jensen et al. (2020) made comparisons among machine learning algorithms for the prediction of invasive Kudzu vine in the USA using S2 and Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) data. They found a higher performance of RF, neural

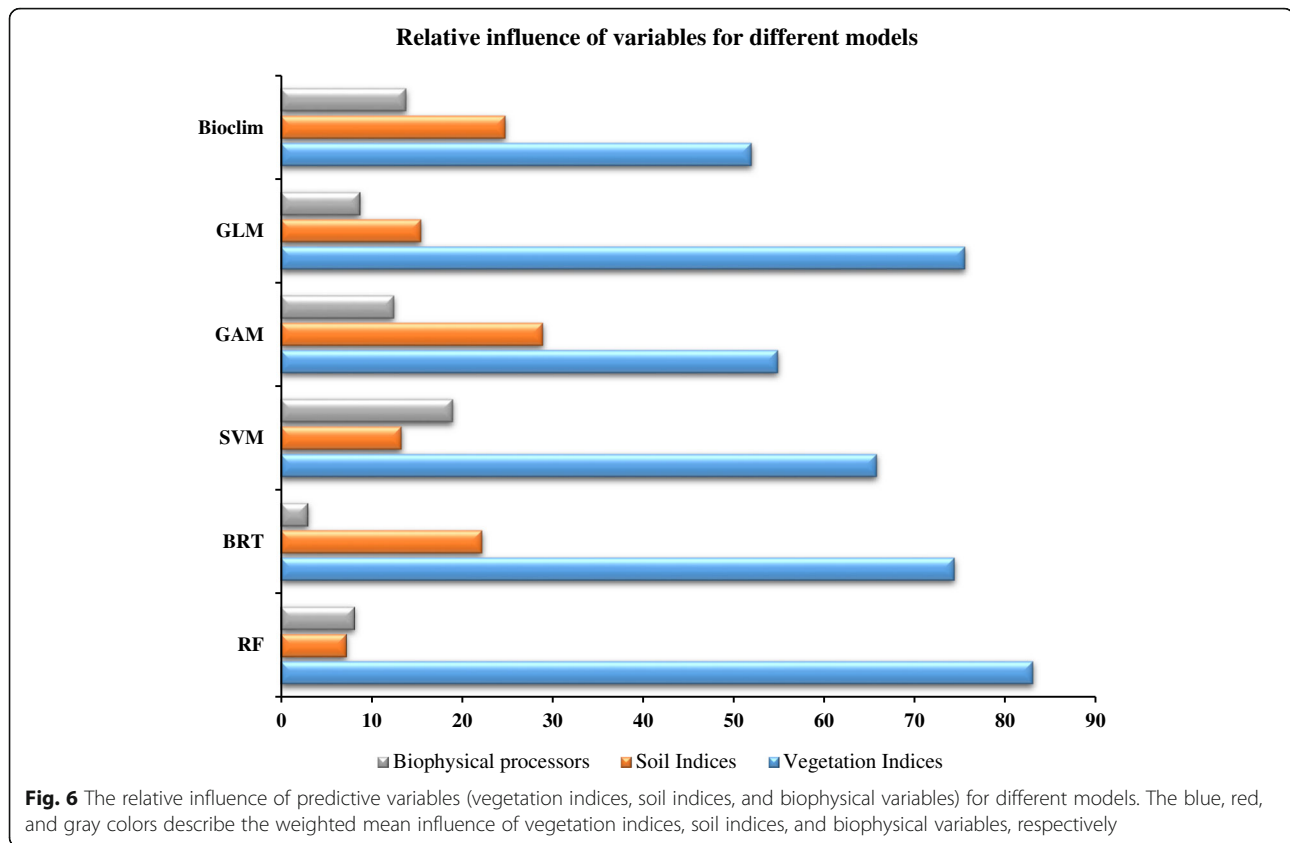




network, and SVM algorithms. Besides, West et al. (2016) investigated five SDMs for the prediction of *Tamarisk* invasive species using remote sensing data and found RF as the best-performed model followed by Maxent, BRT, Multivariate Adaptive Regression Spline (MARS), and GLM. Also, West et al. (2017) compared four SDMs using remote sensing data for the prediction of invasive cheatgrass and concluded that RF was better performed than GLM and BRT. The higher performance of RF is attributed to its ability to avoid overfitting as it combines and votes the most popular class from several individual trees (Breiman 2001). Due to its higher

performance, it is the most flexible and widely applied machine learning algorithm for various field of studies such as land cover classification (Abdi 2020), forest monitoring (Ma et al. 2020), species richness, and density (Kosicki 2020), and invasive species modelling (Ng et al. 2018). In addition, its capability for remote sensing-based studies is also immense as it requires minimum time for satellite image classification (Sabat-Tomala et al. 2020).

Furthermore, the bioclim model was the least performed model in our study and was reaffirmed with the study findings of Hernandez et al. (2006) and Guisan



et al. (2007). Except for bioclim, all models worked well for S2-based studies for the prediction of *Prosopis* distribution and can be used in similar environments.

Potential of Sentinel-2 for invasive species prediction

Our study indicated that mapping and modelling the distribution of invasive *Prosopis* using remote sensing-derived variables have an immense contribution to the management of the invasive species. Our study findings agreed with the government report and previous studies in the area. A study by Shiferaw et al. (2019b) mapped *Prosopis* distribution using Landsat-8 OLI and in situ survey data using RF classifier. They found that 15.4% of the study area was invaded by *Prosopis*. Their result finding was comparable with our findings which were 14.3%.

In addition, our study is also consistent with the study finding of Shiferaw et al. (2019a, b, c) and the report of the Ministry of Livestock and Fisheries (MoLF 2017). Shiferaw et al. (2019a, b, c) found a 12.33% distribution of *Prosopis* using remote sensing, climate, and infrastructural variables collected in 2017. Employing 8.3% as a rate of annual increase Shiferaw et al. (2019b), its distribution will increase to 15.4% of the region in 2020. This result is also comparable to our study findings which are 14.3%. Besides, a report by MoLF (2017) described that

Prosopis covers 12.6% of the Afar region. Considering the same rate of increase (8.3%), its distribution will increase to 15.75% in 2020 which was also comparable to our study findings. On the contrary, a study by Wakie et al. (2014) using remote sensing-derived variables and topo-climatic variables in Afar predicts the distribution of *Prosopis* species and concludes that the distribution of species was 3.8%. Its minimum prediction might occur due to the low spatial resolution of MODIS vegetation indices and bioclimatic variables used for their study. Besides, the quality of WorldClim-based bioclimatic variables is uncertain in areas where weather stations are scarce (Deblauwe et al. 2016).

Martinez et al. (2020) and Truong et al. (2017) described the importance of contemporary remotely sensed variables for the prediction of invasive species. Sentinel-1 and S2 data have a huge contribution in detecting and mapping invasive species (Rajah et al. 2019). The freely available S2 data with its high spatial, spectral, and temporal resolutions provides important information for species-level monitoring and management (Immitzer et al. 2016; Rapinel et al. 2019). Several studies also used S2-derived variables for mapping and modelling of invasive species distribution (Arogoundade et al. 2020; Dube et al. 2020; Ng et al. 2017; Rajah et al. 2019). Dube et al. (2020) compared S2 with Landsat-8 OLI for mapping

the distribution of invasive *Lantana camara* (Verbenaceae) and conclude that the higher performance of S2-derived variables is due to its higher spectral and spatial resolution. Moreover, a study by Arogoundade et al. (2020) applauds the use of S2 spectral bands, S2-derived vegetation indices, and their combination for modelling invasive species. In addition, they also pointed out that red edge S2 bands have a huge contribution to mapping and modelling invasive species distribution.

Among S2-derived variables, our study identified the higher influence of vegetation radiometric indices over soil indices and biophysical variables (Fig. 5). Vegetation indices derived from satellite remote sensing data are recognized as a reliable source of information for monitoring vegetation changes (Feilhauer et al. 2013; He et al. 2015; Maschler et al. 2018; Nouri et al. 2018; Teillet et al. 1997). Several studies also confirmed the higher importance of vegetation indices over other variables such as infrastructure, bio (climate), and remote sensing spectral bands. A study by Shiferaw et al. (2019c) illustrated that the Normalized Difference Vegetation Index (NDVI) had higher importance than other variables for invasive species prediction. Similarly, a study by Wakie et al. (2014) employed remote sensing and topo-climatic variables to map and model invasive *Prosopis* distribution in the Afar region, Ethiopia, and described that the relative influence of Enhanced Vegetation Index (EVI) and NDVI was better than other variables. Immitzer et al. (2019) also demonstrated that S2-derived vegetation indices enhanced model performance.

Furthermore, a study by Rajah et al. (2019) evaluated S2-based Vegetation Indices (S2-VIs) for detecting and modelling invasive American Bramble (*Rubus cuneifolius*) and described that the performance of models with S2-VIs was far better than S2-VIs fused with Sentinel-1 Synthetic Aperture Radar (SAR) and S2 optical imagery. Besides, a study by Arogoundade et al. (2020) evaluated the performance of S2-derived vegetation indices for modelling invasive *Parthenium hysterophorus* in South Africa and confirmed its potential for mapping and modelling invasive species. Moreover, TNDVI and NDVI derived from Landsat-8 OLI showed the higher performances for prediction and description of forest parameters such as density, canopy cover, and basal area (Nouri et al. 2018). Similarly, a study by Musande et al. (2012) evaluated the performance of vegetation indices to discriminate specific crop types and found that the performance of TNDVI was better than other vegetation indices used in their study. However, its limitation to identify water areas and considers it as vegetation cover can significantly decrease the accuracy of the model particularly if the study was conducted in areas where large water bodies are available (Shetty and Somashekar 2013).

However, several studies employed the commonly known vegetation indices, such as NDVI, without evaluating the performances of other variables from vegetation, soil and water radiometric indices, and biophysical variables (e.g., Arogoundade et al. 2020). Considering the different characteristics (benefits and limitations) of radiometric indices and biophysical variables, the use of several indices provides certain predictions as it incorporates different features such as soil, water, and vegetation. Moreover, the performances of radiometric indices varied in different studies. For example, Bannari et al. (2002) described the higher performance of the Transformed Difference Vegetation Index (TDVI) over widely used SAVI and NDVI. They pointed out that TDVI has higher sensitivity to bare soil below vegetation cover that helps to provide necessary information about the specific vegetation parameters. Besides, NDVI has some uncertainties as it is affected by soil reflectance and hence the use of other vegetation indices can reduce this problem (Koller and Upadhyaya 2005).

Besides vegetation indices, the contribution of biophysical variables in monitoring vegetation changes, in our study, was also immense. Mudereri et al. (2019) evaluated the performances of biophysical variables of LAI, FAPAR, CWC, Fraction of Vegetation Cover (FVC), chlorophyll content (Cab), and S2 bands to characterize land cover in semi-arid regions and conclude that both biophysical variables and S2 wavebands have great capability for land cover classification. They also conclude that FAPAR was the best-performing variable and outperforming the other variables used in their study. Also, biophysical variables have a higher capacity for monitoring and managing vegetation changes (Atzberger et al. 2015). Moreover, biophysical variables derived from satellite remote sensing data have a huge contribution to describe forest variables (Schlerf et al. 2005). In particular, S2 data provides an unprecedented option to retrieve biophysical parameters (Brown et al. 2019).

Furthermore, in our study, the relative influence of soil radiometric indices (BI, CI, TSAVI, and RI) was minimum compared to vegetation radiometric indices. According to Nouri et al. (2018), the lower performance of soil radiometric indices was observed in areas where low and high vegetation densities are available. Indeed, low species richness and low diversity of woody vegetation are the major characteristics of our study area (Ilukor et al. 2016). Therefore, evaluating the performance of several indices beyond the commonly used indices is necessary for mapping and modelling invasive species distribution (West et al. 2017). However, in our study, the relative influence of TNDVI was far greater than other vegetation and soil radiometric indices. Hence, our results would have benefited if it includes other bioclimatic variables (Ahmed et al. 2020). However, acquiring

these variables at high resolution was difficult in the study area.

Conclusion

Our study describes the distribution of invasive *Prosopis* in the lower Awash River basin, Ethiopia, using machine learning (RF, BRT, and SVM), regression (GAM, GLM), and profile (bioclim) methods. We used S2-RIs and biophysical variables as predictors to evaluate the performances of models. The performance of machine learning algorithms (RF and BRT) was very high. Besides, the roles of regression models (GAM and GLM) were also found to be very high next to RF, BRT, and ensemble models. On the contrary, the performance of the bioclim model was insufficient. Hence, we encourage researchers not to highly depend on the prediction of the bioclim model with S2-RIs and biophysical variables for predictions of *Prosopis* distribution in the dryland ecosystem. We also encourage researchers to evaluate the performances of models or to use models evaluated previously in related ecosystems and datasets before directly employing specific models as the performances of models can create a significant difference. Therefore, the use of several models can provide reliable information and increase the confidence of ecologists in their result findings.

The best-performed RF model predicted 1354.6 km² (14.3%) of the study area were invaded by the species indicating more efforts are required to reduce its distribution. Our study also demonstrated that the use of freely available S2 data has an immense contribution for detecting, mapping, and modelling the spatial distribution of invasive species with a high level of precision. In particular, the use of S2-RIs and biophysical variables can provide basic information about vegetation, soil, and water for better spatial modelling of invasive species. Also, the higher performances of S2-derived variables for mapping and modelling invasive *Prosopis* distribution indicates the use of such datasets are adequate for such type of studies. Moreover, the relative influences of vegetation radiometric indices were very high followed by soil radiometric indices, biophysical variables, and water radiometric indices. We recommend researchers integrate variables from vegetation indices, soil indices, and biophysical variables for modelling invasive species rather than relying on commonly known vegetation radiometric indices.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13717-021-00285-6>.

Additional file 1 : Table S1. Sentinel-based variables used before multicollinearity analysis: Vegetation Radiometric Indices, Soil radiometric indices, water radiometric indices, and biophysical processors.

Additional file 2 : Table S2. Relative influence of variables (%) for models: Random Forest (RF), Support Vector Machine (SVM), Boosted Regression Trees (BRT), Generalised Additive Model (GAM), Generalized Linear Model (GLM), and bioclim.

Abbreviations

SDM: Species Distribution Modelling; S2-RIs: Sentinel-2 Radiometric Indices; S2: Sentinel 2; TSS: True Skill Statistics; AUC: Area under the curve; RF: Random Forest; GIS: Geographic Information System; NMA: National Metrological Agency; VIF: Variance Inflation Factor; GPS: Global Positioning System; BOA: Bottom Of the Atmosphere; ESA: European Space Agency; SNAP: Sentinel Application; BI: Brightness Index; CI: Colour Index; CWC: Canopy Water Content; FAPAR: Fraction of Absorbed Photosynthetically Active Radiation; LAI: Leaf Area Index; MCARI: Modified Chlorophyll Absorption Ratio Index; MNDWI: Modified Normalized Difference Water Index; MTCI: Meris Terrestrial Chlorophyll Index; RI: Redness Index; S2REP: Sentinel-2 Red-Edge Position Index; TNDVI: Transformed Normalized Difference Vegetation Index; TSAVI: Transformed Soil Adjusted Vegetation Index; BRT: Boosted Regression Trees; SVM: Support Vector Machine; GAM: Generalized Additive Model; GLM: Generalized Linear Model; MCP: Minimum Convex Polygon; ROC: Receiver operator characteristics; GBM: Generalized Boosted regression Models; MARS: Multivariate Adaptive Regression Spline; AVIRIS: Airborne Visible/Infrared Imaging Spectrometer; MODIS: Moderate Resolution Imaging Spectroradiometer; OLI: Operational Land Manager; NDVI: Normalized Difference Vegetation Index; EVI: Enhanced Vegetation Index; SAR: Synthetic Aperture Radar; TDVI: Transformed Difference Vegetation Index; FVC: Fraction of Vegetation Cover; Cab: Chlorophyll content

Acknowledgements

This paper is part of the doctoral study entitled "Role of remote sensing in invasive species distribution modelling, the case of *Prosopis* in the lower Awash River basin, Ethiopia." We would like to thank Wollo University and the Ethiopian Space Science and Technology Institute (ESSTI) for allowing this doctoral study.

Authors' contributions

All authors made a valuable contribution. NA designed and wrote the methodology, collected field data and literature, carried out data analysis, and wrote the draft manuscript; CA refines the methodology, supports in collecting additional literature, reviewed, edited, and rewrite the manuscript; and WZ supports in refining methodology and collection of additional literature coordinated fieldwork and field expanse and refining the manuscript. All authors agreed on the final draft.

Funding

None.

Availability of data and materials

The datasets used and analyzed during the current study are available from the corresponding author.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Remote Sensing, Entoto Observatory and Research Center, Ethiopian Space Science and Technology Institute (ESSTI), Addis Ababa, Ethiopia. ²Department of Geography and Environmental Studies, Wollo University, Dessie, Ethiopia. ³Institute of Geomatics, University of Natural Resources and Life Sciences, Vienna (BOKU), Peter Jordan Straße 82, 1190 Vienna, Austria.

Received: 2 November 2020 Accepted: 4 February 2021

Published online: 06 March 2021

References

- Abdi AM (2020) Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data. *GISci Remote Sens* 57(1):1–20. <https://doi.org/10.1080/15481603.2019.1650447>
- Abdulhafedh A (2017) A novel hybrid method for measuring the spatial autocorrelation of vehicular crashes: combining Moran's Index and Getis-Ord G_i^* Statistic. *Open J Civil Eng* 7:208–221. <https://doi.org/10.4236/ojce.2017.72013>
- Ahmed N, Atzberger C, Zewdie W (2020) Integration of remote sensing and bioclimatic data for prediction of invasive species distribution in data-poor regions: a review on challenges and opportunities. *Environ Syst Res* 9:32
- Allouche O, Tsoar A, Kadmon R (2006) Assessing the accuracy of species distribution models: prevalence, kappa, and the true skill statistic (TSS). *J Appl Ecol* 43:1223–1232. <https://doi.org/10.1111/j.1365-2664.2006.01214.x>
- Arogoundade AM, Odindi J, Mutanga O (2020) Modelling *Parthenium hysterophorus* invasion in KwaZulu-Natal province using remotely sensed data and environmental variables. *Geocarto Int* 35:1450–1465. <https://doi.org/10.1080/10106049.2019.1581268>
- Atzberger C, Darvishzadeh R, Immitzer M, Schlerf M, Skidmore A, Le Maire G (2015) Comparative analysis of different retrieval methods for mapping grassland leaf area index using airborne imaging spectroscopy. *Int J Appl Earth Observ Geoinf* 43(S1):19–31. <https://doi.org/10.1016/j.jag.2015.01.009>
- Ayanu Y, Jentsch A, Müller-Mahn D, Rettberg S, Romankiewicz C, Koellner T (2014) Ecosystem engineer unleashed: *Prosopis juliflora* threatening ecosystem services? *Reg Environ Change* 15(1):155–167. <https://doi.org/10.1007/s10113-014-0616-x>
- Bannari A, Asalhi H, Teillet PM (2002) Transformed difference vegetation index (TDVI) for vegetation cover mapping. In: IEEE International Geoscience and Remote Sensing Symposium. <https://doi.org/10.1109/igarss.2002.1026867>
- Baret F, Guyot G (1991) Potentials and limits of vegetation indices for LAI and APAR assessment. *Remote Sens Environ* 35:161–173. <https://doi.org/10.2136/sssaj1977.03615995004100040037x>
- Benito BM, Cayuela L, Albuquerque FS (2013) The impact of modelling choices in the predictive performance of richness maps derived from species-distribution models: guidelines to build better diversity models. *Methods Ecol Evol* 4:327–335. <https://doi.org/10.1111/2041-210x.12022>
- Bhattacharya M (2013) Machine learning for bioclimatic modelling. *Int J Adv Comput Sci Appl* 4(2):1–8. <https://doi.org/10.14569/ijacsa.2013.04020>
- Booth TH, Nix HA, Busby JR, Hutchinson MF (2014) Bioclim: the first species distribution modelling package, its early applications and relevance to most current MaxEnt studies. *Divers Distrib* 20:1–9. <https://doi.org/10.1111/ddi.12144>
- Bradley BA (2014) Remote detection of invasive plants: a review of spectral, textural, and phenological approaches. *Biol Invasions* 16:1411–1425. <https://doi.org/10.1007/s10530-013-0578-9>
- Breiman L (2001) Random forests. *Machine Learn* 45:5–32. <https://doi.org/10.1201/9780367816377-11>
- Brown LA, Ogutu BO, Dash J (2019) Estimating forest leaf area index and canopy chlorophyll content with Sentinel-2: an evaluation of two hybrid retrieval algorithms. *Remote Sens* 11:1752
- Cernicharo J, Verger A, Camacho F (2013) Empirical and physical estimation of Canopy Water Content: from CHRIS/PROBA Data. *Remote Sens* 5:5265–5284. <https://doi.org/10.3390/rs5105265>
- Cortes C, Vapnik V (1995) Support-vector networks. *Machine Learn* 20:273–297. <https://doi.org/10.1109/64.163674>
- Dash J, Curran PJ (2004) The MERIS terrestrial chlorophyll index. *Int J Remote Sens* 25(23):5403–5413. <https://doi.org/10.1080/0143116042000274015>
- Daughtry CS, Walthall CL, Kim MS, de Colstoun EB, McMurtrey JE (2000) Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote Sens Environ* 74:229–239. <https://doi.org/10.3184/174751911X556684>
- Deblauwe V, Droissart V, Bose R, Sonké B, Blach-Overgaard A, Svenning J-C et al (2016) Remotely sensed temperature and precipitation data improve species distribution modeling in the tropics. *Glob Ecol Biogeogr* 25:443–454. <https://doi.org/10.1111/geb.12426>
- Dube T, Shoko C, Sibanda M, Madileng P, Maluleke XG, Mokwatedi VR et al (2020) Remote sensing of invasive *Lantana camara* (Verbenaceae) in semi-arid savanna rangeland ecosystems of South Africa. *Rangel Ecol Manage* 73:411–419. <https://doi.org/10.1016/j.rama.2020.01.003>
- Elith J, Graha CH, Anderson RP, Dudik M, Ferrier S, Guisan A et al (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29:129–151. <https://doi.org/10.1111/j.1432-1033.1987.tb13499.x>
- Elith J, Leathwick JR (2009) Species distribution models: ecological explanation and prediction across space and time. *Annu Rev Ecol Syst* 40:677–697. <https://doi.org/10.1146/annurev.ecolsys.110308.120159>
- Elith J, Leathwick JR, Hastie T (2008) A working guide to boosted regression trees. *J Anim Ecol* 77:802–813. <https://doi.org/10.1111/j.1365-2656.2008.01390.x>
- Engler R, Waser LT, Zimmermann NE, Schaub M, Berdos S, Ginzler C, Psomas A (2013) Combining ensemble modeling and remote sensing for mapping individual tree species at high spatial resolution. *For Ecol Manage* 310:64–73. <https://doi.org/10.1016/j.foreco.2013.07.059>
- Escadafal R (1989) Remote sensing of arid soil surface color with Landsat thematic mapper. *Adv Space Res* 9(1):159–163. [https://doi.org/10.1016/0273-1177\(89\)90481-X](https://doi.org/10.1016/0273-1177(89)90481-X)
- Evangelista PH, Kumar S, Stohlgren TJ, Jarnevich CS, Crall AW, Norman JB, Barnett DT (2008) Modelling invasion for a habitat generalist and a specialist plant species. *Divers Distrib* 14:808–817. <https://doi.org/10.1111/j.1472-4642.2008.00486.x>
- Feilhauer H, He KS, Rocchini D (2012) Modeling species distribution using niche-based proxies derived from composite bioclimatic variables and MODIS NDVI. *Remote Sens* 4:2057–2075. <https://doi.org/10.3390/rs4072057>
- Feilhauer H, Thonfeld F, Faude U, He KS, Rocchini D, Schmidtlein S (2013) Assessing floristic composition with multispectral sensors — a comparison based on monotemporal and multiseasonal field spectra. *Int J Appl Earth Observ Geoinf* 21:218–229. <https://doi.org/10.1016/j.jag.2012.09.002>
- Fensholt R, Sandholt I, Rasmussen MS (2004) Evaluation of MODIS LAI, fAPAR, and the relation between fAPAR and NDVI in a semi-arid environment using in situ measurements. *Remote Sens Environ* 91:490–507. <https://doi.org/10.1016/j.rse.2004.04.009>
- Fischer D, Thomas SM, Suk JE, Sudre B, Hess A, Tjaden NB et al (2013) Climate change effects on Chikungunya transmission in Europe: geospatial analysis of vector's climatic suitability and virus' temperature requirements. *Int J Health Geogr* 12:51. <https://doi.org/10.1186/1476-072X-12-51>
- Früh L, Kampen H, Kerkow A, Schaub GA, Walther D, Wieland R (2018) Modelling the potential distribution of an invasive mosquito species: comparative evaluation of four machine learning methods and their combinations. *Ecol Model* 388:136–144. <https://doi.org/10.1016/j.ecolmodel.2018.08.011>
- Godoy O, de Lemos-Filho JP, Valladares F (2011) Invasive species can handle higher leaf temperatures under water stress than Mediterranean natives. *Environ Exp Bot* 71:207–214. <https://doi.org/10.1016/j.envexpbot.2010.12.001>
- González-Ferreras AM, Barquín J, Peñas FJ (2016) Integration of habitat models to predict fish distributions in several watersheds of Northern Spain. *J Appl Ichthyol* 32:204–216. <https://doi.org/10.1111/jai.13024>
- Gu H, Wang J, Ma L, Shang Z, Zhang Q (2019) Insights into the BRT (Boosted Regression Trees) method in the study of the climate-growth relationship of Masson pine in subtropical China. *Forests* 10:228. <https://doi.org/10.3390/f10030228>
- Guisan A, Edwards TC, Hastie T (2002) Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecol Model* 157: 89–100. [https://doi.org/10.1016/s0304-3800\(02\)00204-1](https://doi.org/10.1016/s0304-3800(02)00204-1)
- Guisan A, Zimmermann NE, Elith J, Graham CH, Phillips S, Peterson AT (2007) What matters for predicting the occurrences of trees: techniques, data, or species' characteristics? *Ecol Monogr* 77(4):615–630. <https://doi.org/10.1890/06-1060.1>
- Harrell FE, Lee KL, Mark DB (1996) Tutorial in biostatistics multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 15:361–387. [https://doi.org/10.1002/0470023678.ch2b\(i\)](https://doi.org/10.1002/0470023678.ch2b(i))
- He KS, Bradley BA, Cord AF, Rocchini D, Tuanmu M, Schmidtlein S, Turner W, Wegmann M, Pettorelli N (2015) Will remote sensing shape the next generation of species distribution models? *Remote Sens Ecol Conserv* 1:4–18. <https://doi.org/10.1002/rse2.7>
- Hernandez PA, Graham CH, Master LL, Albert DL (2006) The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography* 29:773–785
- Hijmans RJ, Elith J (2019) Spatial distribution models. <https://rsatial.org/sdm/SDM.pdf>. Accessed date 10 June 2020

- Howard PL (2019) Human adaptation to invasive species: a conceptual framework based on a case study meta-synthesis. *Ambio* 48:1401–1430. <https://doi.org/10.1007/s13280-019-01297-5>
- Ilukor J, Birner R, Tilahun M, Getu S (2014) A social-economic assessment of the impact of *Prosopis juliflora* invasion and participative management approach in the Afar Region, Ethiopia. In: *Managing Prosopis juliflora for Better (Agro-) Pastoral Livelihoods in the Horn of Africa*, pp 30–38
- Ilukor J, Rettberg S, Treyde A, Birner R (2016) To eradicate or not to eradicate? Recommendations on *Prosopis juliflora* management in Afar, Ethiopia, from an interdisciplinary perspective. *Pastoralism* 6:14. <https://doi.org/10.1186/s13570-016-0061-1>
- Immitzer M, Neuwirth M, Böck S, Brenner H, Vuolo F, Atzberger C (2019) Optimal input features for tree species classification in Central Europe Based on multi-temporal Sentinel-2 data. *Remote Sens* 11:2599. <https://doi.org/10.3390/rs11222599>
- Immitzer M, Vuolo F, Atzberger C (2016) First experience with Sentinel-2 data for crop and tree species classifications in Central Europe. *Remote Sens* 8:166. <https://doi.org/10.3390/rs8030166>
- Jensen T, Hass FS, Akbar MS, Petersen PH, Arsanjani JJ (2020) Employing machine learning for detection of invasive species using Sentinel-2 and AVIRIS data: the case of *Kudzu* in the United States. *Sustainability* 12:3544. <https://doi.org/10.3390/SU12093544>
- Jiménez-Valverde A (2014) Threshold-dependence as a desirable attribute for discrimination assessment: implications for the evaluation of species distribution models. *Biodivers Conserv* 23:369–385. <https://doi.org/10.1007/s10531-013-0606-1>
- Koller M, Upadhyaya SK (2005) Relationship between modified normalized difference vegetation index and leaf area index for processing tomatoes. *Appl Eng Agric* 21(5):927–934
- Kosicki JZ (2020) Generalised Additive Models and Random Forest Approach as effective methods for predictive species density and functional species richness. *Environ Ecol Stat* 27:273–292. <https://doi.org/10.1007/s10651-020-00445-5>
- Landmann T, Dubovyk O, Ghazaryan G, Kimani J, Abdel-Rahman E (2020) Wide-area invasive species propagation mapping is possible using phenometric trends. *ISPRS J Photogrammetry Remote Sens* 159:1–12. <https://doi.org/10.1016/j.isprsjprs.2019.10.016>
- Lemke D, Brown JA (2012) Habitat modeling of alien plant species at varying levels of occupancy. *Forests* 3(3):799–817. <https://doi.org/10.3390/f3030799>
- Lima TA, Beuchle R, Langner A, Grecchi RC, Griess VC, Achard F (2019) Comparing Sentinel-2 MSI and Landsat 8 OLI imagery for monitoring selective logging in the Brazilian Amazon. *Remote Sens* 11:961. <https://doi.org/10.3390/rs11080961>
- Linders T, Bekele K, Schaffner U, Allan E, Alamirew T, Choge S (2020) The impact of invasive species on social-ecological systems: relating supply and use of selected provisioning ecosystem services. *Ecosyst Serv* 41:101055. <https://doi.org/10.1016/j.ecoser.2019.101055>
- Liu C, Berry PM, Dawson TP, Pearson RG (2005) Selecting thresholds of occurrence in the prediction of species distributions. *Ecography* 28:385–393. <https://doi.org/10.1111/j.0906-7590.2005.03957.x>
- Lopatín J, Dolos K, Hernández HJ, Galleguillos M, Fassnacht FE (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. <https://doi.org/10.1016/j.rse.2015.11.029>
- Ma W, Feng Z, Cheng Z, Chen S, Wang F (2020) Identifying forest fire driving factors and related impacts in China using random forest algorithm. *Forests* 11:507. <https://doi.org/10.3390/F11050507>
- Martinez B, Reaser JK, Dehgan A, Zamft B, Baisch D, McCormick C et al (2020) Technology innovation: advancing capacities for the early detection of and rapid response to invasive species. *Biol Invasions* 22:75–100. <https://doi.org/10.1007/s10530-019-02146-y>
- Maschler J, Atzberger C, Immitzer M (2018) Individual tree crown segmentation and classification of 13 tree species using airborne hyperspectral data. *Remote Sens* 10:1218. <https://doi.org/10.3390/rs10081218>
- Mathieu R, Pouget M, Cervelle B, Escadafal R (1998) Relationships between satellite-based radiometric indices simulated using laboratory reflectance data and typical soil color of an arid environment. *Remote Sens Environ* 66: 17–28. [https://doi.org/10.1016/S0034-4257\(98\)00030-3](https://doi.org/10.1016/S0034-4257(98)00030-3)
- Mehari ZH (2015) The invasion of *Prosopis juliflora* and Afar pastoral livelihoods in the Middle Awash area of Ethiopia. *Ecol Process* 4:13. <https://doi.org/10.1186/s13717-015-0039-8>
- Mendes P, Velazco SJE, Andrade AFA, De Marco P (2020) Dealing with overprediction in species distribution models: how adding distance constraints can improve model accuracy. *Ecol Model* 431:109180. <https://doi.org/10.1016/j.ecolmodel.2020.109180>
- Meroni M, Ng W-T, Rembold F, Leonardi U, Atzberger C, Gadain H, Shaiye M (2017) Mapping *Prosopis juliflora* in West Somaliland with Landsat 8 satellite imagery and ground information. *Land Degrad Develop* 28:494–506. <https://doi.org/10.1002/ldr.2611>
- Mi C, Huettmann F, Guo Y, Han X, Wen L (2017) Why choose Random Forest to predict rare species distribution with few samples in large undersampled areas? Three Asian crane species models provide supporting evidence. *PeerJ* 5:e2849. <https://doi.org/10.7717/peerj.2849>
- MoLF (2017) Federal Democratic Republic of Ethiopia Ministry of Livestock and Fisheries. In: *National Strategy on Prosopis juliflora management* <http://woodyweeds.org/wp-content/uploads/2017/03/Prosopis-Strategy-Ethiopia.pdf>. Accessed date 10 June 2020
- Mudereri BT, Chitata T, Mukanga C, Mupfiga ET, Gwahirira C, Dube T (2019) Can biophysical parameters derived from Sentinel-2 space-borne sensor improve land cover characterization in semi-arid regions? *Geocarto Int*. <https://doi.org/10.1080/10106049.2019.1695956>
- Mulugeta S, Fedler C, Ayana M (2019) Analysis of long-term trends of annual and seasonal rainfall in the Awash River Basin, Ethiopia. *Water* 11:1498. <https://doi.org/10.3390/w11071498>
- Musande V, Kumar A, Kale K, Roy PS (2012) Temporal indices data for specific crop discrimination using fuzzy-based noise classifier. *Int Arch Photogrammetry Remote Sens Spatial Inf Sci* 39-B8:289–294
- Naimi B, Araújo MB (2016) Sdm: a reproducible and extensible R platform for species distribution modelling. *Ecography* 39:368–375. <https://doi.org/10.1111/ecog.01881>
- 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 Observ Geoinf* 53: 76–89. <https://doi.org/10.1016/j.jag.2016.07.019>
- Ng W-T, Rima P, Einzmann K, Immitzer M, Atzberger C, Eckert S (2017) Assessing the potential of Sentinel-2 and Pleiades data for the detection of *Prosopis* and *Vachellia* spp. in Kenya. *Remote Sens* 9:74. <https://doi.org/10.3390/rs9010074>
- Ng W-T, Silva ACO, Rima P, Atzberger C, Immitzer M (2018) Ensemble approach for potential habitat mapping of invasive *Prosopis* spp. in Turkana, Kenya. *Ecol Evol* 8(23):11921–11931. <https://doi.org/10.1002/ece3.4649>
- NMA (2020) National Metrological Agency. <http://www.ethiomet.gov.et/>. Accessed date 10 June 2020
- Nouri A, Kiani B, Hakimi MH, Mokhtari MH (2018) Estimating oak forest parameters in the western mountains of Iran using satellite-based vegetation indices. *J For Res* 31:541–552. <https://doi.org/10.1007/s11676-018-0821-4>
- Paz-Kagan T, Silver M, Panov N, Karnieli A (2019) Multispectral approach for identifying invasive plant species based on flowering phenology characteristics. *Remote Sens* 11:953. <https://doi.org/10.3390/rs11080953>
- Pyšek P, Jarošík V, Hulme PE, Pergl J, Hejd M, Schaffner U, Vilà M (2012) A global assessment of invasive plant impacts on recipient species, communities and ecosystems: the interaction of impact measures, invading species' traits and environment. *Glob Chang Biol* 18(5):1725–1737. <https://doi.org/10.1111/j.1365-2486.2011.02636.x>
- Rajah P, Odindi J, Mutanga O, Kiala Z (2019) The utility of Sentinel-2 Vegetation Indices (VIs) and Sentinel-1 Synthetic Aperture Radar (SAR) for invasive alien species detection and mapping. *Nat Conserv* 35:41–61. <https://doi.org/10.3897/natureconservation.35.29588>
- Rapinel S, Mony C, Lecoq L, Clément B, Thomas A, Hubert-Moy L (2019) Evaluation of Sentinel-2 time-series for mapping floodplain grassland plant communities. *Remote Sens Environ* 223:115–129. <https://doi.org/10.1016/j.rse.2019.01.018>
- Ravindra K, Rattan P, Mor S, Aggarwal AN (2019) Generalized additive models: building evidence of air pollution, climate change and human health. *Environ Int* 132:104987. <https://doi.org/10.1016/j.envint.2019.104987>
- Reaser JK, Burgiel SW, Kirkey J, Brantle KA, Veatch SD, Burgos-Rodríguez J (2020) The early detection of and rapid response (EDRR) to invasive species: a conceptual framework and federal capacities assessment. *Biol Invasions* 22:1–19. <https://doi.org/10.1007/s10530-019-02156-w>
- Reiss H, Cunze S, König K, Neumann H, Kröncke I (2011) Species distribution modelling of marine benthos: a North Sea case study. *Mar Ecol Prog Ser* 442: 71–86. <https://doi.org/10.3354/meps09391>

- Rembold F, Leonardi U, Ng W-T, Gadain H, Meroni M, Atzberger C (2015) Mapping areas invaded by *Prosopis juliflora* in Somaliland on Landsat 8 imagery. *Remote Sens Agric Ecosyst Hydrol XVII* 9637:963723–963721. <https://doi.org/10.1117/12.2193133>
- Reside AE, Watson I, VanDerWal J, Kutt AS (2011) Incorporating low-resolution historic species location data decreases performance of distribution models. *Ecol Model* 222:3444–3448. <https://doi.org/10.1016/j.ecolmodel.2011.06.015>
- Sabat-Tomala A, Raczko E, Zagajewski B (2020) Comparison of support vector machine and random forest algorithms for invasive and expansive species classification using airborne hyperspectral data. *Remote Sens* 12:516. <https://doi.org/10.3390/rs12030516>
- Schlerf M, Atzberger C, Hill J (2005) Remote sensing of forest biophysical variables using HyMap imaging spectrometer data. *Remote Sens Environ* 95: 177–194. <https://doi.org/10.1016/j.rse.2004.12.016>
- Shackleton RT, Le Maitre DC, Pasiecznik NM, Richardson DM (2014) *Prosopis*: a global assessment of the biogeography, benefits, impacts and management of one of the world's worst woody invasive plant taxa. *AoB Plants* 6:plu027. <https://doi.org/10.1093/aobpla/plu027>
- Shetty M, Somashekar R (2013) Multispectral monitoring of vegetation cover of Bangalore metropolitan area. *Glob J Bio-Sci Biotechnol* 2(1):27–32
- Shiferaw H, Schaffner U, Bewket W, Alamirew T, Zeleke G, Teketay D, Eckert S (2019a) Modelling the current fractional cover of an invasive alien plant and drivers of its invasion in a dryland ecosystem. *Sci Rep* 9:1576. <https://doi.org/10.1038/s41598-018-36587-7>
- Shiferaw H, Bewket W, Alamirew T, Zeleke G, Teketay D, Bekele K, Schaffner U, Eckert S (2019b) Implications of land use/land cover dynamics and *Prosopis* invasion on ecosystem service values in Afar Region, Ethiopia. *Sci Total Environ* 675:354–366. <https://doi.org/10.1016/j.scitotenv.2019.04.220>
- Shiferaw H, Bewket W, Eckert S (2019c) Performances of machine learning algorithms for mapping fractional cover of an invasive plant species in a dryland ecosystem. *Ecol Evol* 9:2562–2574. <https://doi.org/10.1002/ece3.4919>
- Soultan A, Safi K (2017) The interplay of various sources of noise on reliability of species distribution models hinges on ecological specialization. *PLoS One* 12(11):e0187906. <https://doi.org/10.1371/journal.pone.0187906>
- Stohlgren TJ, Ma P, Kumar S, Rocca M, Morissette JT, Jarnevich CS, Benson N (2010) Ensemble habitat mapping of invasive plant species. *Risk Anal* 30(2): 224–235. <https://doi.org/10.1111/j.1539-6924.2009.01343.x>
- Szantoi Z, Strobl P (2019) Copernicus Sentinel-2 calibration and validation. *Eur J Remote Sens* 52(1):253–255. <https://doi.org/10.1080/22797254.2019.1582840>
- Tadesse MT, Kumar L, Koech R, Zemadim B (2019) Hydro-climatic variability: a characterization and trend study of the Awash River Basin, Ethiopia. *Hydrology* 6:35. <https://doi.org/10.3390/hydrology6020035>
- Teillet PM, Staenz K, Williams DJ (1997) Effects of spectral, spatial, and radiometric characteristics on remote sensing vegetation indices of forested regions. *Remote Sens Environ* 61:139–149
- Tilahun M, Birner R, Ilukor J (2017) Household-level preferences for mitigation of *Prosopis juliflora* invasion in the Afar region of Ethiopia: a contingent valuation. *J Environ Plann Manag* 60(2):282–308. <https://doi.org/10.1080/09640568.2016.1152955>
- Truong TTA, Hardy G, Andrew ME (2017) Contemporary remotely sensed data products refine invasive plants risk mapping in data-poor regions. *Front Plant Sci* 8:770. <https://doi.org/10.3389/fpls.2017.00770>
- Vilà M, Espinar JL, Hejda M, Hulme PE, Jarošík V, Maron JL et al (2011) Ecological impacts of invasive alien plants: a meta-analysis of their effects on species, communities, and ecosystems. *Ecol Lett* 14:702–708. <https://doi.org/10.1111/j.1461-0248.2011.01628.x>
- Vuolo F, Zoltak M, Pipitone C, Zappa L, Wenng H, Immitzer M et al (2016) Data service platform for Sentinel-2 surface reflectance and value-added products: system use and examples. *Remote Sens* 8:938. <https://doi.org/10.3390/rs8110938>
- Wakie T, Evangelista P, Laituri M (2012) Utilization assessment of *Prosopis juliflora* in Afar Region, Ethiopia. US Forest Service, USDA Office of International Programs, USAID Pastoral Livelihoods Initiative II Project (PLI II), pp 1–15 http://www.nrel.colostate.edu/projects/csu-ethiopia/documents/NewsFeed_Wakie_Final.pdf. Accessed date 10 June 2020
- Wakie TT, Evangelista PH, Jarnevich CS, Laituri M (2014) Mapping current and potential distribution of non-native *Prosopis juliflora* in the Afar region of Ethiopia. *PLoS One* 9(11):e112854. <https://doi.org/10.1371/journal.pone.0112854>
- West AM, Evangelista PH, Jarnevich CS, Kumar S, Swallow A, Luizza MW, Chignell SM (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 Observ Geoinf* 59:135–146. <https://doi.org/10.1016/j.jag.2017.03.009>
- West AM, Evangelista PH, Jarnevich CS, Young NE, Stohlgren TJ, Talbert C, Talbert M, Morissette J, Anderson R (2016) Integrating remote sensing with species distribution models; mapping tamarisk invasions using the software for assisted habitat modeling (SAHM). *J Vis Exp* 116:e54578. <https://doi.org/10.3791/54578>
- Wisn MS, Hijmans RJ, Li J, Peterson AT, Graham CH, Guisan A et al (2008) Effects of sample size on the performance of species distribution models. *Divers Distrib* 14:763–773. <https://doi.org/10.1111/j.1472-4642.2008.00482.x>
- Xu H (2006) Modification of normalized difference water index (NDWI) to enhance open water features in remotely sensed imagery. *Int J Remote Sens* 27(14):3025–3033. <https://doi.org/10.1080/01431160600589179>
- Xu Z (2014) Potential distribution of invasive alien species in the upper Ili river basin: determination and mechanism of bioclimatic variables under climate change. *Environ Earth Sci* 73:779–786. <https://doi.org/10.1007/s12665-014-3083-2>
- Zhang X, Friedl MA, Schaap CB, Strahler AH, Hodges JCF, Gao F (2003) Monitoring vegetation phenology using MODIS. *Remote Sens Environ* 84:471–475. [https://doi.org/10.1016/S0034-4257\(02\)00135-9](https://doi.org/10.1016/S0034-4257(02)00135-9)
- Zimmermann NE, Edwards TC, Moisen GG, Frescino TS, Blackard JA (2007) Remote sensing-based predictors improve distribution models of rare, early successional and broadleaf tree species in Utah. *J Appl Ecol* 44(5):1057–1067. <https://doi.org/10.1111/j.1365-2664.2007.01348.x>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)