GISCIENCE & REMOTE SENSING
2021, VOL. 58, NO. 4, 483–500
https://doi.org/10.1080/15481603.2021.1903281

Modeling the geographic spread and proliferation of invasive alien plants (IAPs) into new ecosystems using multi-source data and multiple predictive models in the Heuningnes catchment, South Africa

Bhongolethu Mtengwana [a], Timothy Dube [a], Bester Tawona Mudereri [a,b] and Cletah Shoko [c]

[a]Department of Earth Sciences, University of the Western Cape, Bellville, South Africa; [b]Department of Animal and Wildlife Science, Midlands State University, Gweru, Zimbabwe; [c]Division of Geography, School of Geography, Archaeology and Environmental Studies, University of Witwatersrand, Johannesburg, South Africa

ABSTRACT

The geographic spread and proliferation of Invasive Alien Plants (IAPs) into new ecosystems requires accurate, constant, and frequent monitoring particularly under the changing climate to ensure the integrity and resilience of affected as well as vulnerable ecosystems. This study thus aimed to understand the distribution and shifts of IAPs and the factors influencing such distribution at the catchment scale to minimize their risks and impacts through effective management. Three machine learning Species Distribution Modeling (SDM) techniques, namely, Random Forest (RF), Maximum Entropy (MaxEnt), Boosted Regression Trees (BRT) and their respective ensemble model were used to predict the potential distribution of IAPs within the catchment. The current and future bioclimatic variables, environmental and Sentinel-2 Multispectral Instrument satellite data were used to fit the models to predict areas at risk of IAPs invasions in the Heuningnes catchment, South Africa. The present and two future climatic scenarios from the Community Climate System Model (CCSM4) were considered in modeling the potential distribution of these species. The two future scenarios represented the minimum and maximum atmospheric carbon Representative Concentration Pathways (RCP) 2.6 and 8.5 for 2050 (average for 2041–2060). The results show that IAPs are predicted to expand under the influence of climate change in the catchment. Concurrently, riparian zones, bare areas, and the native vegetation which is rich in biodiversity will greatly be affected. The mean diurnal range (Bio2), warmest quarter maximum temperature (Bio5), and the warmest quarter precipitation (Bio18) were the most important bioclimatic variables in modeling the spatial distribution of IAPs in the catchment. Comparatively, all the models were successful in predicting the potential distribution of IAPs for all the scenarios. The BRT, MaxEnt, and RF predicted the spatial distribution of IAPs with an Area Under Curve (AUC) of 0.89, 0.92, and 0.94, respectively. The study highlighted the importance of multi-source data and multiple predictive models in predicting the current and potential future IAP distribution. The results from this study provide baseline information for effective land management, planning, and continuous monitoring of the further spread of IAPs within the Heuningnes catchment.

ARTICLE HISTORY
Received 12 September 2020
Accepted 8 March 2021

KEYWORDS
Biological invasion risk; climate change; ensemble; ecological niche modeling; remotely sensed data

1. Introduction

Globally, Invasive Alien Plants (IAPs) often outcompete the native species, which leads to their extinction and reduced biodiversity (Mooney 2005; Wilcove et al. 1998). The establishment and success of IAPs into new ecosystems is mainly caused by environmental changes because of anthropogenic influences and climate change (Buckley, Catford, and Gibson 2016). It is expected that the increase in temperatures may facilitate and accelerate the spread of IAPs while reducing the resilience of natural vegetation (Ncube et al. 2020; Tarabon et al. 2018). This will likely increase the areas at risk of invasion due to the increased competition, hence causing massive losses in biodiversity as a result of species range shifts or extinctions. These changes are a great concern to the conservation and preservation of native species, water resources, and biodiversity management.

Currently, the low precipitation and the already warmer temperatures in Africa make the continent to be more vulnerable to the expected extreme climate change conditions such as the vulnerability to the impacts of IAPs (IPCC 2014; Kotir 2010). The future climate in Africa is likely to experience temperature

CONTACT Bhongolethu Mtengwana 3544906@myuwc.ac.za.

increases between 3°C and 6°C before the end of the century (Serdeczny et al. 2016). Therefore, many regions in Southern Africa will experience sharp increases in temperatures and frequent droughts (IPCC 2014). These changes will likely trigger mass extinctions due to the loss of the biological conditions suitable for most indigenous species resulting in the opportunistic spread of IAPs. This will be due to the increased invasibility of host ecosystems as a result of extreme climatic events (Masters and Norgrove 2010). Lazzaro et al. (2020) highlighted the impacts of IAPs on natural plant communities mainly because of competition. It further looked at the need for strategies to overcome the impacts of IAPs on indigenous species, particularly in the light of climate change effects. Therefore, there is a great and urgent need to accurately model and predict the potential current and future distributions of IAPs to empirically prioritize areas for control, mitigation, and adaptation.

Localized modeling of IAPs provides critical insights into the processes driving vegetation dynamics, community structure, and the general functioning of ecosystems, including anticipated impacts. It has been shown that the use of machine learning algorithms to model species distributions for alien and native species provides useful information on the response of vegetation species to climate change (Ndlovu et al. 2018; Ncube et al. 2020). For instance, De La Hoz et al. (2019) and Hoveka et al. (2016) observed that some plant species will decrease in extent while others increase because of climate change. In a different study, Vorsino et al. (2014) also reported the vulnerability of ecosystems to climate change. The study further indicated that IAPs spatial distribution dynamics can be successfully determined using Species Distribution Modeling techniques (SDM) and machine learning (ML) approaches. ML models such as the Random Forest (RF), Maximum entropy (MaxEnt), and Boosted Regression Trees (BRT) using 'presence-only' data have proved their robustness and their ability to produce good predictive performances as well as their versatility to handle autocorrelations (Crase et al. 2012; Fourcade et al. 2014). The BRT model uses a maximum likelihood approach to merge multiple models to improve on a single regression tree (Elith et al. 2018). In the RF model, the prediction is produced by selecting the class with the highest random combinations in multiple decision trees (Bangira et al. 2019). MaxEnt predicts the species occurrence by finding the maximum entropy of the spatial distribution, i.e. largest spread (Merow et al. 2013). These machine learning tools have been widely used for modeling species distributions in ecological applications. For instance, the use of RF in SDM has been demonstrated by several studies (Mudereri et al. 2020a; Zhang et al. 2019). Yu, Cooper, and Infante (2020) demonstrated the ability to use BRT in improving the predictive ability for species distribution modeling. However, these models often produce slightly differing predictive results because of their different algorithmic architecture and input data assumptions. To resolve the resulting uncertainties, the ensemble modeling approach used in this current study is relatively popular because of its ability to combine multiple models' predictive strengths and reduce their individual weaknesses (Hao et al. 2019; Naimi et al. 2014; Ng et al. 2018; Stohlgren et al. 2010). This increases the predictive modeling capabilities of the individual models (Mudereri et al. 2020a; Ng et al. 2018). For instance, the ensemble of RF and MaxEnt was preferred for mapping the distribution of alien *Chromolaena odorata* and *Mikania micrantha* to reduce spatial uncertainties of the predictions due to their reported performance in a study conducted by Nath et al. (2019). Also, Ng et al. (2018) modeled the invasive *Prosopis* species using multiple SDMs and concluded that individual SDMs achieved high accuracies while the ensemble model achieved the highest scores.

While it is common practice to exclusively use the bioclimatic predictors, incorporating remotely sensed data and environmental variables such as topography, land cover, and other geographical ancillary data has been reported by earlier studies to improve the predictive ability of models (Truong, Hardy, and Andrew 2017; Vorsino et al. 2014; West et al. 2017). However, to the best of the authors' knowledge, only a limited number of studies aimed at predicting the occurrence IAPs using data that combines computer-generated bioclimatic data (current and future), remotely sensed data, and environmental variables in the global south and particularly in Sub-Saharan Africa. Therefore, coupling the Sentinel-2 multispectral data that has strategically placed bands and more vegetation-sensitive bands, with other environmental variables, has the potential to increase species discrimination and improve the performance of the prediction and mapping. Several studies have already demonstrated that adding remotely sensed data from Sentinel-2 improves modeling, classification, and predictions (Forkuor et al. 2017; Malahlela, Adjorlolo, and Olwoch 2019; Mudereri et al. 2019; Ndlovu et al. 2018).

Therefore, in this study, we explored the use of multi-source data *viz.* bioclimatic, topographic, and Sentinel-2 data as predictor variables in predicting the distribution of IAPs in varying climatic scenarios. This was aimed to improve the understanding of the potential impacts of IAPs at the catchment scale using MaxEnt, RF, BRT, and their respective ensemble model. Additionally, the study sought to establish the key climatic factors and their influence on IAP distribution under current and projected future climatic conditions. Modeling IAPs under different projected climate scenarios allows better evaluation and anticipation of future changes in distribution, thus providing empirical and effective management and control (Landmann et al. 2020). Predicting the distribution of these species under extreme future climate conditions using Representative Concentration Pathways (RCP 2.6 and RCP 8.5) for best-case and worst-case scenarios will provide insights into the behavior of these species under these extremes.

2. Materials and methods

2.1 Study area

The Heuningnes catchment is situated within the Overberg region in the province of the Western Cape, South Africa (Figure 1). It lies between the latitudes 34°19` S and 34°50` S and longitudes of 19° 35` E and 20°18` E covering a relatively small area of approximately 1 442 km^2. Elevation ranges in the study area are between the sea level and ~837 m a.s.l. The area experiences a Mediterranean climate characterized by hot dry summers (November to March) with the maximum temperature of up to 27° C and wet cold winters (May to August) with minimum temperatures below 10°C (Mkunyana et al. 2018). The average annual rainfall in the catchment is 500 mm/year where most of the rainfall occurs in the mountainous region of the catchment and fed

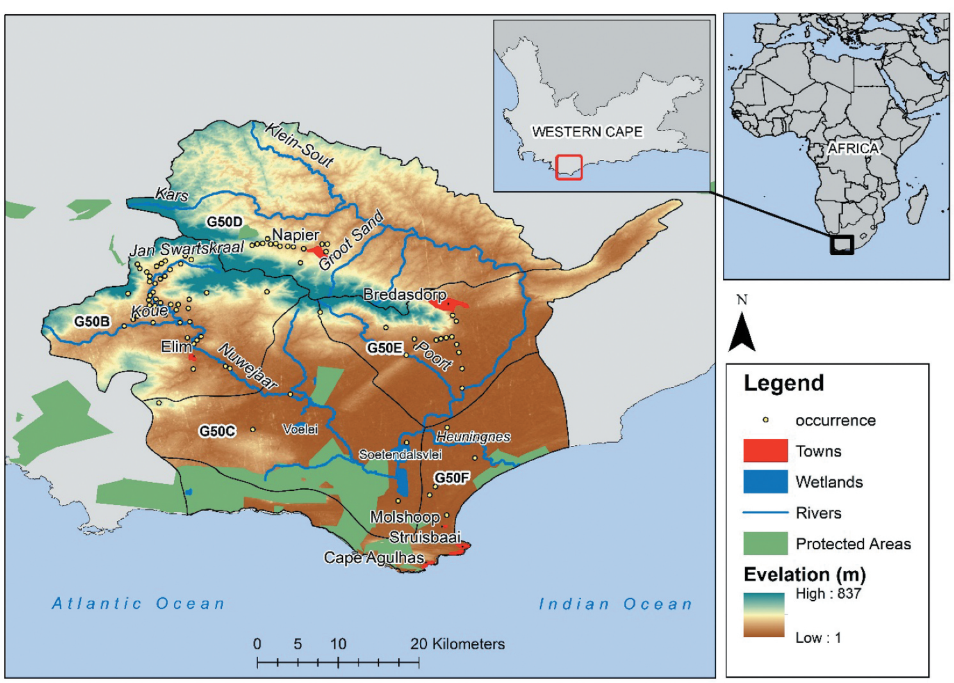

Figure 1. The study area showing five sub-catchments (G50B – G50F).

down streams to low-lying areas where complex wetland systems occur. Crop cultivation is one of the major land use activities with very little urban development taking place. This catchment falls within the Cape Agulhas in the Cape Floristic Region characterized by endemic species comprising fynbos as the main indigenous vegetation often within a restricted range. The region has the greatest proportion of land invaded by IAPs according to the survey done by Kotzé et al. (2010) and Le Maitre, Versfeld, and Chapman (2000). These species were initially introduced in the area for timber, windbreaks, and stabilization of sand dunes and have since become widespread. The dominant *Acacia* species found in the catchment pose threats to the biodiversity, water resources, protected areas, grazing lands, and the ecosystem at large. The occurrence of these dominant IAPs in the catchment is largely found along the Nuwerjaar and upstream areas mostly within riparian and adjacent to the mountainous region. Their aggressive spread especially within riparian zones has led to the establishment of a forum to coordinate and implement the clearing of IAPs continuously. The frequent and continuous removal and burning of these species in attempts to mitigate their spread and thus impacts depict its ability to spread rapidly. Therefore, monitoring the spread of IAPs in this catchment, and understanding climatic conditions which may influence its spread is essential.

2.2 Reference field data

A total number of 244 ‘presence-only’ occurrence data of the IAPs were used to model the potential distributions of IAPs in the catchment. The field survey to identify IAPs was conducted in August 2018. These reference data were used for the three SDMs for predicting the potential species distribution and habitat suitability. The reference data were collected using a purposive sampling approach that targeted areas of dense IAPs stands identified in accessible site areas. Each of the sampling units was approximately 30 m x 30 m in dimension. The points were collected at the approximate center of each of the dense IAPs stands to eliminate the edge-effect. For each IAP stand identified, a handheld Garmin eTrex Global Positioning System (GPS) was used to record the reference of the occurrence points at an error margin of ±3 m.

2.3 Predictor variables

2.3.1 Sentinel-2 data acquisition and pre-processing

The processing level-1 C Sentinel-2 data of the 24th of August 2018 was obtained from the USGS Earth Explorer platform (http:/earthexplorer.usgs.gov) in three granules (T34HCG, T34HDG, and T34HCH). These tiles were mosaicked into a single scene that covered the entire study area. The satellite image acquisition date was selected to align with the field data collection period for the day of low cloud cover (<5%) and availability from the sensor archive. This processing level (1 C) is provided as Top of the Atmosphere (TOA) reflectance, which has been orthorectified in cartographic geometry in tiles of 100 km^2 and projected to the UTM/WGS84 Zone 35S projection. Atmospheric correction was performed using the Sen2Cor processor with default settings in SNAP software version 6.0. The bands that were

Table 1. Spectral and spatial characteristics of the Sentinel-2 data that were considered in modeling IAPs distribution in the catchment with bold showing fitting variable(s).

Band name	Band number	Band center (nm)	Pixel size (resolution)	Potential application
Blue	B2	490	10	Atmosphere
Green	B3	560	10	Vegetation
Red	B4	665	10	Vegetation
Red-edge (RE1)	B5	705	20	Vegetation
Red-edge (RE2)	B6	740	20	Vegetation
Red-edge (RE3)	B7	783	20	Vegetation
Near-infrared (NIR)	**B8**	**842**	**10**	**Vegetation**
Narrow near-infrared (NIRn)	B8a	865	20	Vegetation
Short wave infrared	B11	1610	20	Vegetation
Short wave infrared	B12	2190	20	Vegetation

Table 2. The environmental, bioclimatic, and terrain variables considered to predict IAPs distribution. The predictor variables in bold were selected for the final modeling after removing highly correlated variables.

Environmental variable description	Bioclim code	Unit
Annual mean temperature	Bio1	°C
Mean diurnal range	**Bio2**	**°C**
Iso-thermality	Bio3	-
Temperature seasonality	Bio4	-
Maximum temperature of the warmest month	**Bio5**	**°C**
Minimum temperature of the coldest month	**Bio6**	**°C**
Temperature annual range	Bio7	°C
Mean temperature of wettest quarter	Bio8	°C
Mean temperature of driest quarter	Bio9	°C
Mean temperature of warmest quarter	Bio10	°C
Mean temperature of coldest quarter	Bio11	°C
Annual precipitation	Bio12	mm
Precipitation of wettest month	**Bio13**	**mm**
Precipitation of driest month	Bio14	mm
Precipitation seasonality	Bio15	-
Precipitation of wettest quarter	Bio16	mm
Precipitation of driest quarter	**Bio17**	**mm**
Precipitation of warmest quarter	**Bio18**	**mm**
Precipitation of coldest quarter	Bio19	mm
The direction of the slope	**Aspect**	-
Altitude above sea level	Elevation	m
Angle of inclination	**Slope**	**degrees**
Topographic index	**TPI**	-
Moisture index	**TWI**	-
Thematic land cover classes	**Land cover**	-
Soil characteristics	**Soil types**	-

considered in modeling the distribution of IAPs are indicated in Table 1. These bands have also been used by other studies to classify major land cover classes, including IAPs with an overall accuracy of >70% (Ncube et al. 2020; Ndlovu et al. 2018). The data were resampled to the 30 m pixel size with the bioclimatic and environmental variables. These reflectance bands were used as inputs into the model, and to produce a land cover distribution showing major land cover types within the catchment.

2.3.2 Topographic data

The details of topographic variables considered in predicting the distribution of IAPs are presented in Table 2. Digital Elevation Model (DEM: https://dwtkns.com/srtm30m/) of 30 m spatial resolution was used as the elevation variable and to generate aspect, slope, Topographic Wetness Index (TWI), and Topographic Position Index (TPI) for the study area. Aspect and slope were generated from the DEM using the QGIS terrain analysis plugin (QGIS Development Team 2019). Terrain variables influence soil type, soil moisture, sun angle, precipitation hence the distribution of vegetation (Bennie et al. 2006; Perring 1956, 1959). The used soil type data was retrieved from the ISRIC data hub (http://data.isric.org/). TWI is an index for soil moisture which affects vegetation growth and composition (Gábor et al. 2020). TWI has also been successfully used for studying vegetation patterns and predicting the spatial distribution of plants (Sørensen et al. 2006). The TWI was derived based on equation 1:

$$TWI = \ln\left(\frac{a}{\tan\beta}\right) \qquad (1)$$

where a is the local upslope area and $\tan\beta$ is the slope (Beven and Kirkby 1979)

TPI is generally used to categorize landform types in an area and describes the biophysical processes occurring on landscapes, which can be key in predicting habitat suitability and species distribution (Seif 2014; Weiss 2001). It is defined as the difference between the elevation of a cell in a DEM and a mean elevation of neighboring cells. Equation 2 shows the calculation of TPI.

$$TPI = M_0 - \sum_{n-1} M_n/_n \qquad (2)$$

where M_0 is the elevation of the DEM point being evaluated, M_n is the elevation of the pixel grid, and n is the total sum of the surrounding points (Mokarram, Roshan, and Negahban 2015).

2.3.3 Bioclimatic data

Bioclimatic data have been widely used in SDMs to determine and explain factors driving species distributions (Booth 2018; Gallardo et al. 2017; Ndlovu et al. 2018). The bioclimatic data are generated from monthly rainfall and temperatures. These data sets can be used to explain the potential species distributions by providing biologically meaningful variables that convey annual and seasonal mean climate conditions as well as intra-year seasonality (Hijmans et al. 2005; O'donnell and Ignizio 2012). A total number of 19 bioclimatic variables (Table 2) representing each scenario for the current (1950–2000) and future climate (2050) were freely obtained from WorldClim (http://www.worldclim.org/) at 30 arc seconds spatial resolution (~1 km x 1 km). The obtained future climate scenarios were based on the fourth Community Climate System Model (CCSM4) projections (Gent et al. 2011; Mohammadi et al. 2019). Only two of the four atmospheric carbon Representative Concentration Pathways (RCPs) namely RCP 2.6 (minimum emission) and RCP 8.5 (maximum emission) proposed by the Intergovernmental Panel on Climate Change (IPCC) were selected to show the possible minimum and maximum impacts respectively. The RCP scenarios represent the minimum and maximum radioactive forces of 2.6 and 8.5 watts/m^2 for the CO_2 concentrations by 2050 (IPCC 2014)

The future bioclimatic variables based on the minimum and maximum RCPs for temperature and precipitation were used to determine how the projected climate changes will vary to the current climate. Additionally, the change in the most important bioclimatic variables was also calculated. This was achieved by subtracting the projected climatic conditions of the variables from the current climatic conditions following Ncube et al. (2020). The objective was to show the relative increase or decrease in the projected climate to determine how the variations affect the predicted distribution of IAPs within the study area.

2.4 Collinearity test for the bioclimatic variables

The problem associated with multicollinearity between predictor variables in SDMs is the inflation of coefficient standard errors, making some variables insignificant or resulting in model overfitting (Akinwande, Dikko, and Samson). The coefficient of Pearson's correlation and the Variance Inflation Factor (VIF) were used to remove highly correlated variables from the models (Akinwande, Dikko, and Samson). The threshold used for the collinearity test for Pearson's correlation was set at $|r|>0.7$ while for VIF it was set to 10 (Dormann et al. 2013; Makori et al. 2017). The VIF measures the degree to which multicollinearity increases the slope estimate variance, based on regressing paired predictor variables against each other in multiple regression (Plant 2012). The 'usdm' package in R-software was used for eliminating variables with high VIF and thus modeling the distribution (Naimi et al. 2014; R Core Team 2019). The threshold was set at $th = 0.7$ where values greater than the threshold are considered to be highly correlated within a model (Dormann et al. 2013; Richard et al. 2018). Therefore, all variables identified as having a high correlation based on the set thresholds were removed.

A total number of 12 variables were selected for the current and future prediction (Table 2). Only the land cover derived from Sentinel-2 satellite bands was eligible for model parameterization excluding the reflectance spectral bands which have been excluded because of high collinearity. All data sets used were projected to the WGS84 coordinate system and clipped to the area of the catchment using the open-source QGIS version 3.8.2 (QGIS Development Team 2019). The selected variables used for final modeling were then resampled to the 30 m pixel size.

2.5 Predicting the distribution of IAPs in Heuningnes catchment

Semiautomatic generation of 1 000 'pseudo-absence' points within the SDM package in R was used together with the collected 'presence-only' occurrence to create a 'presence-background' file. The use of presence-only models with pseudo-absence has been widely applied considering the challenge of obtaining 'absence data' (Downie, Von Numers, and Boström 2013). Only three modeling techniques, namely, the BRT, RF, and MaxEnt were used from the 15 modeling techniques available within the 'sdm' package. The syntax, sample R-code, and the step-by-step description of how to run the SDM package are provided (see Naimi et al. 2014; Naimi and Araújo 2016; Naimi 2020) for reproducible species distribution modeling.

Table 3. R packages and references of the three models used in predicting the IAP distribution.

Model algorithm	'sdm' syntax	Package	Reference
Boosted regression trees	'brt'	'gbm'	(Elith et al. 2008)
Random forest	'rf'	'randomForest'	(Liaw and Wiener 2002)
MaxEnt	'maxent'	'dismo'	(Phillips et al. 2006)
Ensemble	'ensemble'	'sdm'	(Naimi and Araújo 2016)

These models were selected for use in this study because they produce relatively high accurate results and provide predictions within geographically complex environments, such as in our study area (Barakat et al. 2018; Makaya et al. 2019; Mudereri et al. 2019). Table 3 summarizes the relevant functions and packages used in predicting IAPs distribution for the three models.

After removing variables showing high collinearity for the current prediction, all three selected models were fitted and ran using the bioclimatic variables, i.e. Bio2, Bio5, Bio6, Bio13, Bio17, Bio18, and land cover, soil type, aspect, slope, TPI, TWI, and Band 8 of Sentinel 2 data. Similarly, the future predictions were run using the same bioclimatic variables as used for the current prediction and all the qualifying environmental and topographic variables highlighted in Table 2. An ensemble modeling approach was further used to reduce the spatial differences occurring from the predictions of the tested models. Ensemble models fit and maximize the prediction accuracy of the different machine learning approaches using a weighted average of the highest performance from each model (Araújo et al.). Therefore, the weighted average of the TSS (True Skill Statistics) was used to produce the ensemble model since it comparatively improves the predictive power of the model when contrasted with using the mean or median (Naimi and Araújo 2016). The TSS is a widely used threshold-dependent measure of the model performance and reliable measure to combine different models compared to AUC which is highly sensitive to the occurrence of the observations (Allouche et al. ; Richard et al. 2018). The threshold of TSS = 0.7 was set to qualify the models for inclusion in the ensemble. The variable importance values were computed, using the randomization method which computes Pearson's correlation between reference predictions and the shuffled variables, inherent in the 'sdm' package used.

The QGIS software was further used to process the outputs of all three models with their respective ensemble into maps for map-making and quantification of the potential habitat of IAPs in the Heuningnes catchment. The outputs of the three predictive models and their respective ensemble models were used to calculate the suitable areas for the occurrence of IAPs in the form of a binary raster image, i.e. <0.3 unsuitable and ≥0.3 suitable. In each of these suitability categories, the total number of image pixels was then used to estimate the suitability or unsuitability of IAPs coverage within the catchment.

2.6 Model evaluation

Measuring the performance of a model is important to test the accuracy and reliability of its outcomes (Fois et al. 2018). The accuracy of the models was tested, using a 10-fold cross-subsampling approach. The performances of the models were measured using the Area Under Curve (AUC) and TSS (Allouche et al.). The AUC values range between 0 and 1, where inaccurate models have values closer to 0 while models with an AUC value ≥0.7 show high predictive abilities (Mohammadi et al. 2019). On the other hand, the TSS is a product of sensitivity (proportion of true positives) and specificity that explains commission and omission errors performed by a model (Kyalo et al. 2018). Similarly, the range of TSS values is between −1 and +1. The TSS values closer to +1 demonstrate a perfect agreement between the observations and predictions while TSS ≤ 0 indicates no agreement and thus poor modeling performance (Allouche et al. ; Somodi, Lepesi, and Botta-Dukát 2017).

2.7 The general flow of the process used in modeling the IAPs potential distribution

Figure 2 shows the four stages that were considered in modeling the distribution of IAPs and the

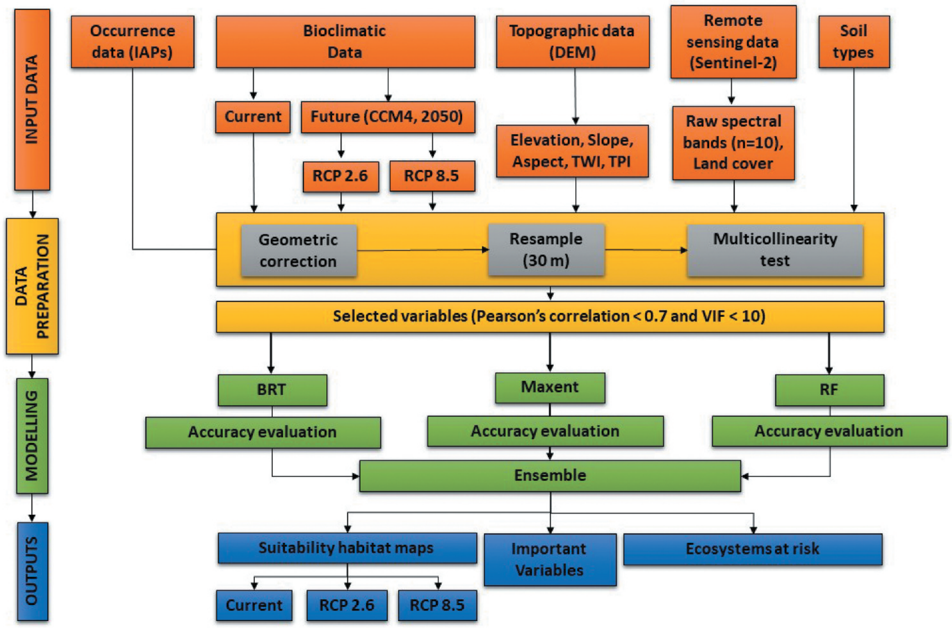

Figure 2. The processes undertaken to determine current and future suitable habitats for IAPs.

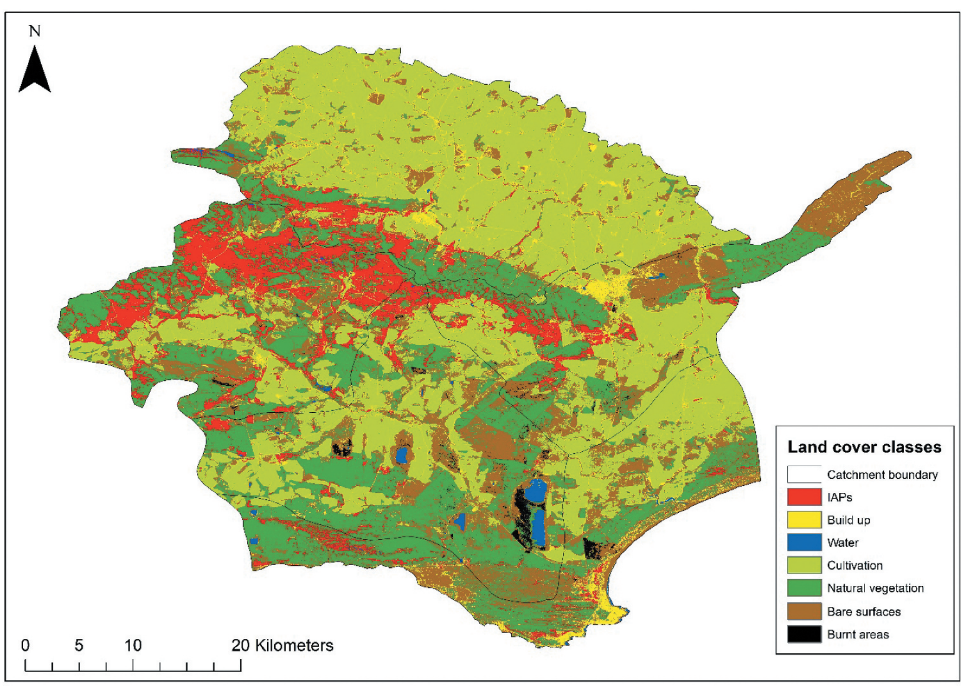

Figure 3. Derived land use map of the catchment using the Sentinel-2 satellite data (Mtengwana et al. 2020).

respective processes undertaken at each modeling stage. The stages included input data which involved data collection and consideration of predictor variables to be included. This was followed by the predictor variable preparation for modeling, using the three selected models and their ensemble. Finally, the important bioclimatic variables were identified, and the outputs of the potentially suitable habitats ensemble were obtained for the climate scenarios. The mapping of risk areas was produced to pinpoint ecosystems most susceptible to the predicted IAPs distribution.

3. Results

3.1 Land use and land cover across the catchment using Sentinel-2 data

Figure 3 shows the distribution of IAPs and other land use and land cover classes across the catchment. The accuracy of the image classification results for the current land cover use yielded an overall accuracy of 71% (see Mtengwana et al. 2020). The most common land use within the catchment is areas under cultivation, particularly in the northern parts. IAPs are predominantly within the central belt, whereas natural vegetation occupies the southern parts of the catchment, with some bare surface areas. Also, among the different quaternary catchments, G50B seems to be the most invaded by IAPs compared to other quaternary catchments. G50D and G50E are greatly characterized by cultivated lands with some extent of invaded areas. G50C is characterized by the occurrence of wetlands of varying sizes.

3.2. Changes in temperature and precipitation due to climate change

Temperature and precipitation are the general factors used to recognize the effects of climate change. From the bioclimatic variables, the calculated changes from the CCMS4 model show that the annual mean temperatures will increase for both RCP 2.6 and RCP 8.5 (Table 4). However, RCP 2.6 has a greater magnitude of increment in annual mean temperature compared to the RCP 8.5 projection. The annual precipitation also shows a general decrease in both future RCPs, with an increase in mean for RCP 8.5. Therefore, the catchment is expected to receive lower rainfall and increased temperatures.

3.3 Model performances for predicted species distribution under current climatic

RF and MaxEnt were moderately constant in their prediction among the replicated models compared to BRT as shown by the produced Receiver Operating Curves (ROC) in Figure 4. The RF model (AUC = 0.93 and TSS = 0.82) yielded the highest accuracy metrics for both AUC and TSS followed by MaxEnt with BRT obtaining the least accuracies. Further, all models show high values of specificity and sensitivity as demonstrated by the high values of TSS produced by both RF and MaxEnt (TSS > 0.8). All reported accuracies are based on current bioclimatic climatic variables. Accuracy was not measured for 2050 variables since there are no reference presence data for the future timestamp period.

Table 4. Projected changes in bioclimatic variables for 2050 in Heuningnes catchment. Increases are shown by the positive values while negative values show decreases by the specified magnitude.

			Changes	
Parameter		Current	RCP 2.6	RCP 8.5
Annual mean temperature (°C)	Min	14.55	1.97	1.55
	Mean	16.83	2.25	1.75
	Max	17.67	2.34	1.84
Annual precipitation (mm)	Min	427	−28.00	−18.00
	Mean	487	−7.00	9.00
	Max	619	−26.00	−4.00

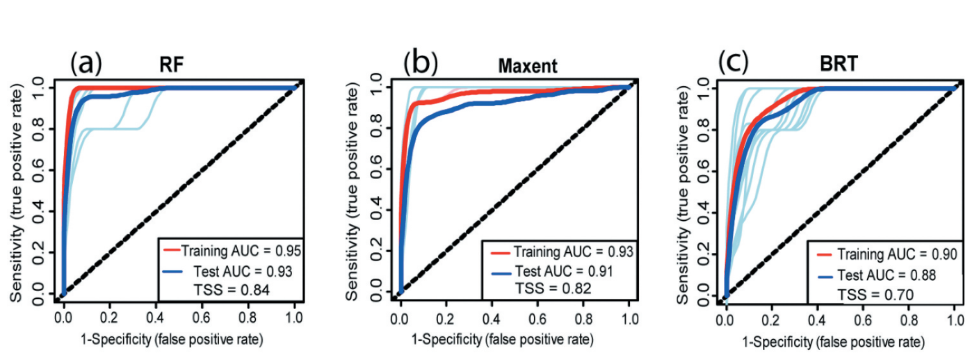

Figure 4. The ROC for (a) RF (b) MaxEnt and (c) BRT. The red curve represents the smoothened mean AUC of the 10-fold cross-validation subsampling (light blue curves) using the training data, while the dark blue curve depicts the mean AUC using the test data.

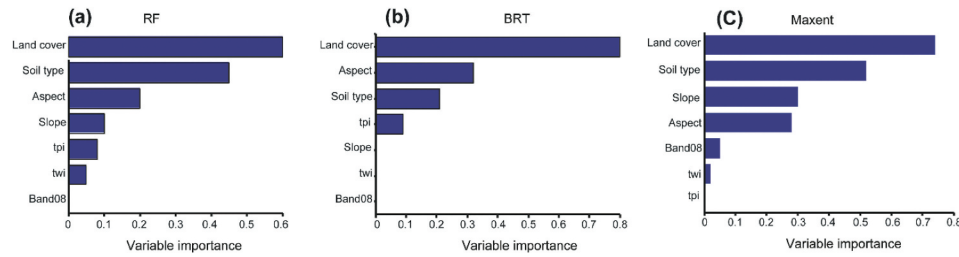

Figure 5. Variable importance measure for the prediction of IAPs using the current climatic scenario.

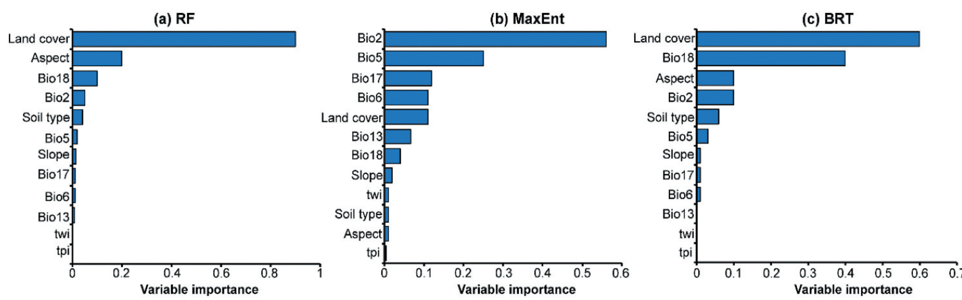

Figure 6. Variable importance measure for the prediction of IAPs under the future climatic scenario.

3.4 The most relevant predictors for the IAP distribution

The most relevant variable for the modeling of the current species distribution across all models was land cover (Figure 5). Soil type was the second important variable in both RF and MaxEnt, with the aspect as the third most important variable, respectively. Sentinel-2, band 8 (NIR centered at 842 nm) was among the least important variables in all three models.

The important variables for future climate were similar for RF and BRT except for aspect and Bio18 with MaxEnt showing different variable importance (Figure 6). The land cover was the most relevant non-climatic predictor across all the models, while TPI was the least important variable. Bio18 and Bio2 were the most important bioclimatic variables for RF and BRT, while for the MaxEnt model, Bio2 and Bio5 were the most dominant. Notably, the variable importance measure for the MaxEnt model was dominated by bioclimatic factors. The variation among the variable importance predictors between the models can be accounted for by the unique statistical approaches of each model. Also, the comparison of these variables across the models shows the influence of climate in predicting species distributions and land cover as a fundamental driver of habitat suitability.

3.5 Prediction of potential distribution

The predicted distributions vary across the models but show a similar pattern with suitable areas mostly occurring in the central regions of the catchment (Figure 7). However, BRT predictions show very distinct spatial differences in the southern part of the catchment when compared to both the MaxEnt and RF in all three climatic scenarios. MaxEnt shows the expansion of IAPs in RCP8.5 while showing a contraction in the RCP2.6 relative to the current prediction. This contraction is also observed in both future climate scenarios in RF. However, the future suitable areas for the occurrence of IAPs show expansion in both RCP2.6 and RCP8.5. This expansion of IAPs is shown to be toward the southeast part of the catchment, along the riparian zones in the G50B sub-catchment, with great intensity. Overall, BRT shows clear spatial differences from the predicted suitable areas detected by MaxEnt and RF SDM models. To counteract the differences caused by the architecture of each of the models, the ensemble modeling approach was used. The ensemble model provides a weighted average of the predictions and eliminates the spatial uncertainty across the models, thus leveraging on the strength of each model and offsetting their weaknesses and limitations.

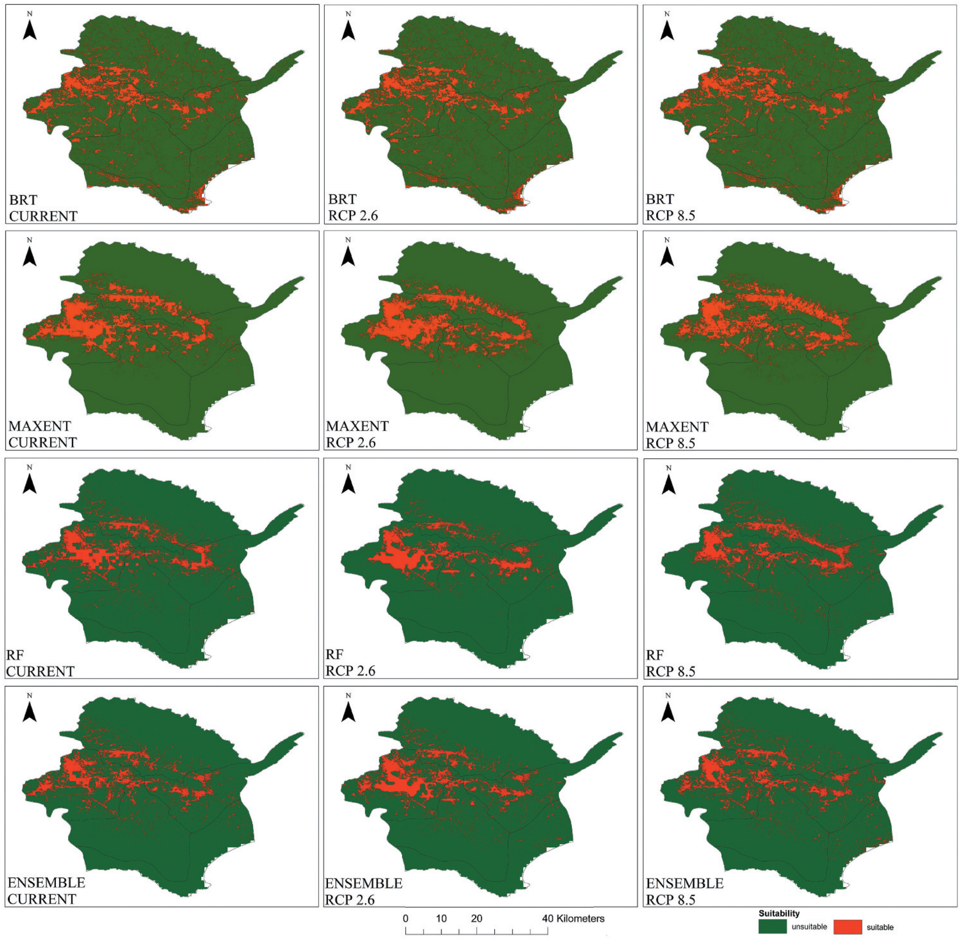

Figure 7. Predicted IAPs suitability maps derived using the three machine learning algorithms used and their respective ensemble. The rows show respective model predictions while the columns present both the current and future climate scenarios. The red areas represent suitable habitats while the green areas signify unsuitable areas.

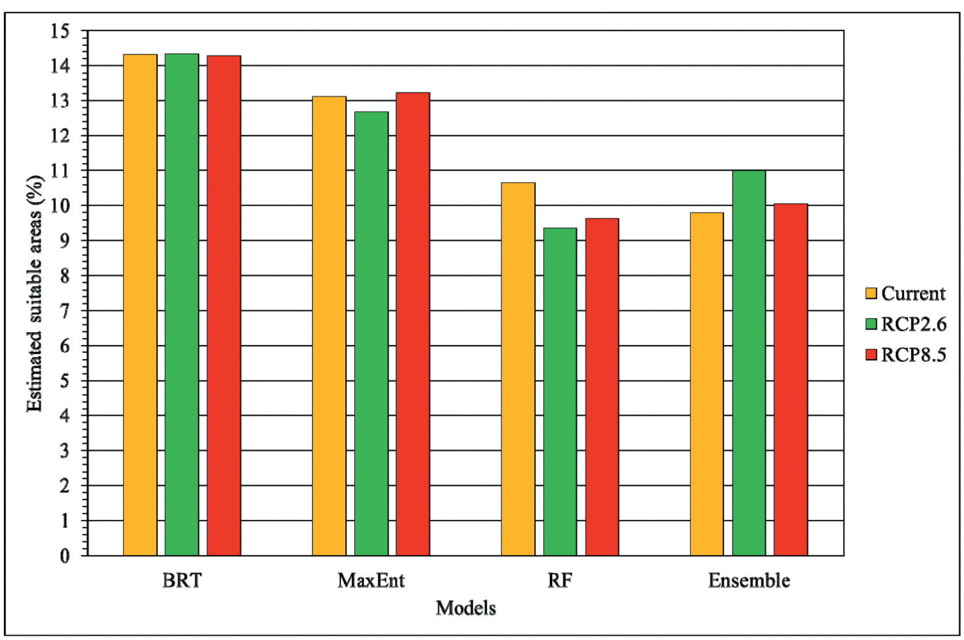

Figure 8. Estimated suitable areas (%) for the occurrence of IAP distribution in Heuningnes catchment for the current and future climate scenarios (RCP 2.6 and RCP 8.5).

3.6 Quantification of the potential habitat of IAPs in Heuningnes catchment

Figure 8 shows the estimated suitable area percentage for the occurrence and distribution of IAPs. The BRT model shows that the estimated areas suitable for IAPs currently is 14.32% and this will increase by 0.01% for RCP 2.6 and decrease to 14.28% in RCP 8.5. For MaxEnt, it is expected that the suitable habitats will decrease to 12.67%, for RCP 2.6, and increase to 13.21% under RCP 8.5 from the current predicted 13.12%. RF shows a decrease from the current 10.64% suitable areas in both RCP 2.6 and RCP 8.5 to 9.97% and 9.63%, respectively. Nevertheless, RCP 2.6 shows a greater decrease than RCP 8.5. Generally, the percentage of the estimated areas varies across all three individual models. However, the overall predictions using the ensemble model show increases in suitability areas for IAPs in both RCP 2.6 and RCP 8.5 by 1.21% and 0.25%, respectively.

3.7 The potential risk of invasion by IAPs in the Heuningnes catchment

The results of the predicted IAPs distribution demonstrate the future invasion range and potential negative impacts, which could result due to the spread of IAPs (Figure 9). It is shown that the currently most infested sub-catchments (G50B, G50D, and G50E) are most vulnerable to the further spread of IAPs. The areas adjacent to the Jan Swartskraal and Koue rivers will likely be greatly affected. These rivers upstream feed lower catchment, and invasion could mean reduced streamflow downstream. The areas adjacent to the major wetlands (Voevlie and Soetendalsvlei) showed some extent of suitable areas, which could potentially invade the wetlands in the future. The areas surrounding the settlements are more susceptible to invasion. The protected areas likely to be considerably invaded are those with already established IAPs; hence, these areas do not show a great extent of susceptibility.

4. Discussion

The continued naturalization and spread of IAPs creates a major concern on how climate change will influence the distribution of these species as climate change is anticipated to modify the dynamics and ecological niches of many species both locally and globally (Lazo-Cancino et al. 2020). As a result, this can even be more detrimental to the ecosystem's provision of services with the impacts severely affecting both biodiversity and hydrological systems (Otieno, Nahrung, and Steinbauer 2019). This study aimed to investigate the likely climate change effects

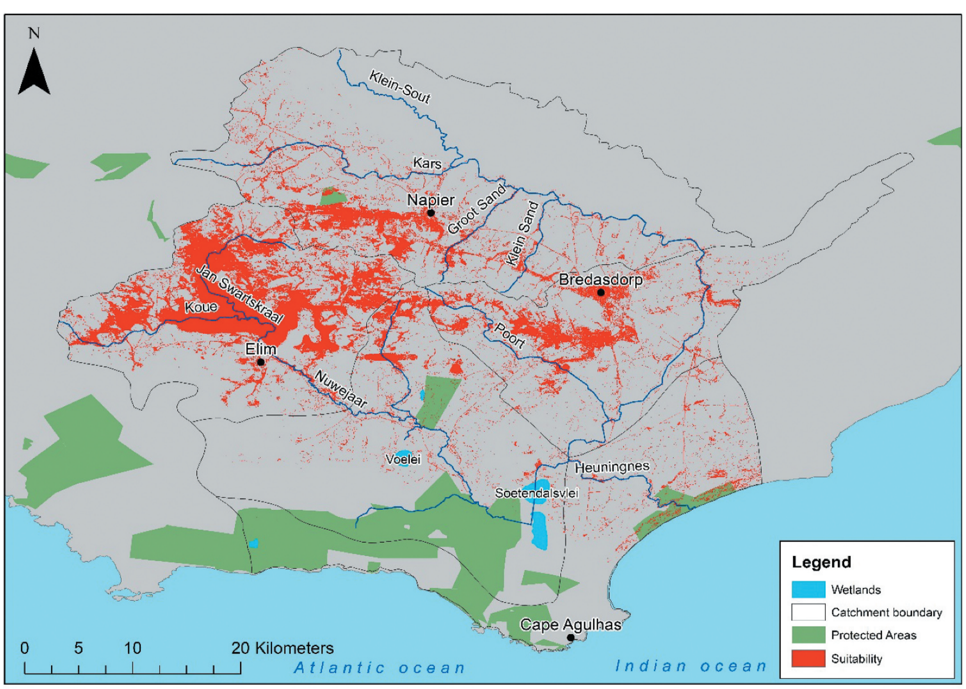

Figure 9. Potential risk area map posed by IAPs in the Heuningnes catchment, using ensemble predictions.

on the distribution of IAPs under the minimum (RCP2.6) and maximum (RCP8.5) climate projections by applying the machine learning approach, using BRT, MaxEnt, RF, and the ensemble. It is imperative to explore different models to identify the models that can accurately predict the species distribution to develop optimized model approaches (Araújo et al. ; Beaumont et al. ; Warren, Matzke, and Iglesias 2019). To achieve this, reputable machine learning algorithms and multisource datasets were successfully used to predict the potentially suitable areas for IAPs at 30 m spatial resolution in Heuningnes catchment, South Africa.

4.1 Predicted and estimated future distribution patterns of IAPs

The overall predicted distribution showed that IAPs abundance will increase toward plains, particularly riparian zones, mostly in sub-catchment 'G50B' where most invasion currently occurs. This was also reported by Kotzé et al. (2010) that *Acacias* are likely to occur within river flood plains. Some parts of the cultivated and naturally vegetated areas also show great suitability for IAPs. Our findings are in line with those of Gutierres et al. (2011) who found that these species can be associated with lowlands, agricultural lands, and margins of lakes in Sesimbra County, Portugal. Furthermore, it was estimated that the suitable potential habitats of IAPs currently cover ~9% of the study area and will increase to ~11%. This increase conveys that suitable habitats have not been fully invaded and will continue under the influence of the changing climate. Notably, it has been pointed out that IAPs have not reached equilibrium in South Africa (Rouget et al. 2004). Despite the high accuracy of our models, the potentially suitable areas for IAPs in this study could have been underestimated due to sampling effort, with predicted suitable areas not showing some of the currently invaded areas and the small difference between the currently invaded areas and future predicted suitable habitats for 2041 to 2060. This is also because the dominant *Acacia* species are known for their rapid spread.

4.2 The most relevant predictor variables

There was a variation in the importance of predictor variables across the models which can be related to the predictive power of the models and their respective underlying algorithms. These observations suggest that the prediction of suitable habitats is dependent on the type of model used since each model employs a different set of equations or algorithms to perform the predictions (Mudereri et al. 2020a). Nonetheless, the land cover showed to be an important predictor variable for IAP distribution in BRT and RF with climate variables showing dominance in the MaxEnt. Although current land cover was established as a very important variable and also used in similar studies such as Ye et al. (2018) and Pang, De Alban, and Webb (2021) for future prediction, the results of the predictions obtained must be used with caution as land cover will likely change in the future.

Nonetheless, other studies have also shown that land cover is an important driver of habitat change (Ndlovu et al. 2018). In contrast, the land cover had minor importance in modeling IAP distribution in a study conducted by Terzano et al. (2018) on a larger scale. On the other hand, some studies (Nath et al. 2019; Terzano et al. 2018) have shown that climate predictors are the most important variables in predicting species distribution; this was partially demonstrated in this study by the MaxEnt model. The incorporation of these important variables, however, has been understood to provide realistic predictions for suitable habits (Thalmann et al. 2015). The mean diurnal range, the maximum temperature of the warmest quarter, and the precipitation of the warmest quarter were the most important bioclimatic predictor variables. Even though remote sensing data facilitates the prediction of IAPs over inaccessible areas (Pearce and Boyce 2006), reflectance spectral bands showed little contribution in the prediction of suitable habitats for IAPs except the land cover derived from these bands. Other studies were able to show relatively considerable contributions of remote sensing derivatives, such as vegetation indices (Mudereri et al. 2019b). Therefore, the use of remotely sensed derived variables, such as vegetation indices, may provide more insights into species physiochemical properties for improved prediction than reflectance spectral bands.

4.3 Impacts of IAPs under the current and projected climate changes

Climate projections suggest potential increases in the annual mean temperature for the catchment, while there will be an observable decrease in annual precipitation. Declines in available water resources and rainfall patterns have already been observed in the study area due to climate variability and drought impacts (Orimoloye et al. 2019). Additionally, the dominant and rapid spreading of *Acacia* species (*A. saligna, A. longifolia* and *A. cyclops*) has been observed in the catchment. These species are likely to adapt to these new anticipated conditions since they show high drought tolerance (Ivanova and Symes 2019). Their increasing spread in riparian zones will largely contribute to reduced streamflow (Prinsloo and Scott 1999). It has been found that these species are most likely dependent on surface water and thus may be a great threat when expanding to these areas (Sher, Wiegand, and Ward 2010). It was also found that the water use of *A. longifolia* occurring in riparian zones in low-lying areas than in hillslopes was dependent on soil moisture and used more water (Mkunyana et al. 2018). Protected areas, natural vegetation, particularly low shrubland (fynbos) are potentially at risk of being invaded causing biodiversity loss due to increased competition for available ecosystem resources. These areas are to a greater extent already invaded by IAPs. Therefore, the predicted future expansion of IAPs will exacerbate the negative impacts on the rivers, wetlands, and biodiversity of the catchment.

4.4 Evaluation of the model performances

The predictions of the potential distribution of IAPs were better than random (AUC and TSS > 0.5) for all three individual models. It was noted that RF produced the highest accuracy followed by MaxEnt and BRT with marginal differences. Similar studies by (Guan et al. 2020) and (Stohlgren et al. 2010) showed the same pattern with the latter models predicting IAPs habitat suitability at relatively high accuracy across the models, although based on different algorithms (Downie, Von Numers, and Boström 2013; Mohammadi et al. 2019; Pearce and Boyce 2006). The robustness of these models was further evident in the spatial distribution of predicted suitable habitats. All three candidate models predicted a similar distribution pattern across all the climatic scenarios in major suitable areas, although spatial differences can also be observed. This could be attributed to the predictive power of the algorithm and the approaches used by each model (Araújo et al.). For example, both MaxEnt and RF models, which performed better than BRT did not predict suitable habitats along the southern catchment boundary in all three climate scenarios. This contradicts the land cover results, which show the presence of IAPs occurrence close to built-up areas in the southernmost part of the catchment. This can suggest a reduced ability to deal with sampling bias toward areas where sampling is most accessible. Even though MaxEnt can handle sparse and irregular occurrence data, it assumes that the area of interest is systematically sampled (Kramer-Schadt et al. 2013).

Several studies showed that there is no convincing evidence to suggest that there is an overall model that is better than all (Guo et al. 2019; Hao et al. 2019; Mudereri et al. 2020b). Therefore, the use of the ensemble analysis becomes paramount in all predictive modeling, especially for producing a realistic and encompassing prediction (Araújo et al.). As such, the ensemble model was successfully used to produce predictions by including only models with a TSS > 0.7 as opposed to AUC due to associated criticisms to ensure only strong models are included (Allouche et al.). The advantage of ensembles is their ability to minimize the spatial uncertainties of the models for each climate scenario to enable reliable spatial estimates (Downie, Von Numers, and Boström 2013; Guan et al. 2020; Pearce and Boyce 2006).

Although the findings of this study provide critical insights on the current and potential distribution and shift of IAPs in the Heuningnes Catchment, South Africa, there is a need for further research to investigate their distribution in detail. For instance, there is a need to map and quantify areas affected by species invasive species in the area. The use of high spatial resolution spatial data like Worldview or Rapid Eye and Unmanned Aerial Vehicles (AUVs) has the potential to identify specific species.

5. Conclusions

Climate change effects characterized by reduced rainfall and increased temperatures will facilitate the distribution of IAPs and increase their abundance in the catchment.

Riparian zones, low-lying areas, and natural shrublands are the most vulnerable areas and must be prioritized in management efforts to reduce the impacts on biodiversity loss and water losses through increased evapotranspiration. These results have also demonstrated the combination of multiple strong predictive models to reduce spatial uncertainties for realistic suitable habitat predictions for effective management practices. The estimated areas suitable for IAPs in this study are better than random but may have been underestimated. Further investigation is required by considering species-specific potential distribution and more ecologically meaningful remotely sensed derived variables as opposed to reflectance spectral bands. Nonetheless, the results provide useful insights into the effective management of IAPs and may be used for prioritized monitoring.

Acknowledgements

This work was financially supported by the Centre of Excellence for Invasion Biology (C·I·B) through Masters research programme and AIMS NEI small grants.

Disclosure statement

All authors declare that they have no conflict of interest.

ORCID

Bhongolethu Mtengwana http://orcid.org/0000-0003-0592-8150
Timothy Dube http://orcid.org/0000-0003-3456-8991
Bester Tawona Mudereri http://orcid.org/0000-0001-9407-7890
Cletah Shoko http://orcid.org/0000-0003-4222-3539

References

Bangira, T., Alfieri, S. M., Menenti, M. and Van Niekerk, A. 2019. "Comparing thresholding with machine learning classifiers for mapping complex water." *Remote Sensing*11(11), 1351. doi:10.3390/rs11111351

Barakat, A., Khellouk, R., El Jazouli, A., Touhami, F. and Nadem, S. 2018. "Monitoring of forest cover dynamics in eastern area of Béni-Mellal Province using ASTER and Sentinel-2A multispectral data." *Geology, Ecology, and Landscapes* 2(3): 203–215. doi:10.1080/24749508.2018

Bennie, J., M. O. Hill, R. Baxter, and B. Huntley. 2006. "Influence Of Slope And Aspect On Long-Term Vegetation Change In British Chalk Grasslands." *Journal of Ecology* 94: 355–368. doi:10.1111/j.1365-2745.2006.01104.x.

Beven, K. J., and M. J. Kirkby. 1979. "A Physically Based, Variable Contributing Area Model Of Basin Hydrology." *Hydrological Sciences Journal* 24: 43–69. doi:10.1080/02626667909491834.

Booth, T. H. 2018. "Why Understanding The Pioneering And Continuing Contributions Of Bioclim To Species Distribution Modelling Is Important." *Austral Ecology* 43: 852–860. doi:10.1111/aec.12628.

Buckley, Y. M., J. Catford, and D. Gibson. 2016. "Does The Biogeographic Origin Of Species Matter? Ecological Effects Of Native And Non-Native Species And The Use Of Origin To Guide Management." *Journal of Ecology* 104: 4–17. doi:10.1111/1365-2745.12501.

Crase, B., Liedloff, A. C. and Wintle, B. A. 2012. "A new method for dealing with residual spatial autocorrelation in species distribution models." *Ecography* 35:879–888. doi:10.1111/j.1600-0587.2011.07138.x

De La Hoz, C. F., E. Ramos, A. Puente, and J. A. Juanes. 2019. "Climate Change Induced Range Shifts In Seaweeds Distributions In Europe." *Marine Environmental Research* 148: 1–11. doi:10.1016/j.marenvres.2019.04.012.

Dormann, C. F., J. Elith, S. Bacher, C. Buchmann, G. Carl, G. Carré, J. R. G. Marquéz, et al. 2013. "Collinearity: A Review Of Methods To Deal With It And A Simulation Study Evaluating Their Performance." *Ecography* 36: 27–46. doi:10.1111/j.1600-0587.2012.07348.x.

Downie, A.-L., M. Von Numers, and C. Boström. 2013. "Influence Of Model Selection On The Predicted Distribution Of The Seagrass Zostera Marina." *Estuarine, Coastal and Shelf Science* 121–122: 8–19. doi:10.1016/j.ecss.2012.12.020.

Elith, J., Leathwick, J. R. and Hastie, T. 2008. "A working guide to boosted regression trees." *Journal of Animal Ecology*77(4): 802–813. doi:10.1111/j.1365-2656.2008.01390.x

Fois, M., Cuena-Lombraña, A., Fenu, G. and Bacchetta, G. 2018. "Using species distribution models at local scale to guide the search of poorly known species: Review, methodological issues and future directions." *Ecological Modelling* 385: 124–132. doi:10.1016/j.ecolmodel.2018.07.018

Forkuor, G., K. Dimobe, I. Serme, and J. E. Tondoh. 2017. "Landsat-8 Vs. Sentinel-2: Examining the Added Value of Sentinel-2's Red-Edge Bands to Land-Use and Land-Cover Mapping in Burkina Faso." *GIScience & Remote Sensing* 55: 331–354. doi:10.1080/15481603.2017.1370169.

Fourcade, Y., Engler, J.O., Rödder, D. and Secondi, J. 2014. "Mapping species distributions with MAXENT using a geographically biased sample of presence data: a performance assessment of methods for correcting sampling bias." *PloS one*9(5): p.e97122. doi:10.1371/journal.pone.0097122

Gábor, L., Moudrý, V., Lecours, V., Malavasi, M., Barták, V., Fogl, M., Šímová, P., Rocchini, D. and Václavík, T. 2020. "The effect of positional error on fine scale species distribution models increases for specialist species." *Ecography* 43(2): 256–269. doi:10.1111/ecog.04687

Gallardo, B., D. C. Aldridge, P. Gonzalez-Moreno, J. Pergl, M. Pizarro, P. Pysek, W. Thuiller, C. Yesson, and M. Vila. 2017. "Protected Areas Offer Refuge From Invasive Species Spreading Under Climate Change." *Global Change Biology* 23: 5331–5343. doi:10.1111/gcb.13798.

Gent, P. R., G. Danabasoglu, L. J. Donner, M. M. Holland, E. C. Hunke, S. R. Jayne, D. M. Lawrence, et al. 2011. "The Community Climate System Model Version 4." *Journal of Climate* 24: 4973–4991. doi:10.1175/2011JCLI4083.1.

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

Guo, Y., X. Li, Z. Zhao, and Z. Nawaz. 2019. "Predicting The Impacts Of Climate Change, Soils And Vegetation Types On The Geographic Distribution Of *Polyporus Umbellatus* In China." *Science Of The Total Environment* 648: 1. doi:10.1016/j.scitotenv.2018.07.465.

Gutierres, F., A. Gil, E. Reis, A. Lobo, C. Neto, H. Calado, and J. C. Costa. 2011. "*Acacia Saligna* (Labill.) H. Wendl In The Sesimbra County Invaded Habitats And Potential Distribution Modeling." *Journal of Coastal Research* SI 64 (Proceedings of the 11th International Coastal Symposium), Szczecin, Poland, ISSN 0749-0208.

Hao, T., J. Elith, G. Guillera-Arroita, J. J. Lahoz-Monfort, and J. Serra-Diaz. 2019. "A Review Of Evidence About Use And Performance Of Species Distribution Modelling Ensembles Like BIOMOD." *Diversity and Distributions* 25: 839–852. doi:10.1111/ddi.12892.

Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. "Very High Resolution Interpolated Climate Surfaces For Global Land Areas." *International Journal of Climatology* 25: 1965–1978. doi:10.1002/joc.1276.

Hoveka, L. N., B. S. Bezeng, K. Yessoufou, J. S. Boatwright, and M. Van Der Bank. 2016. "Effects Of Climate Change On The Future Distributions Of The Top Five Freshwater Invasive Plants In South Africa." *South African Journal of Botany* 102: 33–38. doi:10.1016/j.sajb.2015.07.017.

IPCC 2014. Climate Change 2014. "Synthesis Report. In: Versión Inglés, Climate Change 2014: Synthesis Report. Contribution Of Working Groups I, II And III To The Fifth Assessment Report Of The Intergovernmental Panel On Climate Change".

Ivanova, I. M. and Symes, C. T. 2019. "Invasion of Psittacula krameri in Gauteng, South Africa: are other birds impacted?" *Biodiversity and Conservation* 28(13): 3633–3656.

Kotir, J. H. 2010. "Climate Change And Variability In Sub-Saharan Africa: A Review Of Current And Future Trends And Impacts On Agriculture And Food Security." *Environment, Development and Sustainability* 13: 587–605. doi:10.1007/s10668-010-9278-0.

Kotzé, J. D. F., B. H. Beukes, E. C. Van Den Berg, and T. S. Newby 2010. "National Invasive Alien Plant Survey. Agricultural Research Council – Institute For Soil, Climate And Water".

Kramer-Schadt, S., J. Niedballa, J. D. Pilgrim, B. Schröder, J. Lindenborn, V. Reinfelder, M. Stillfried, I. Heckmann, A. K. Scharf, and D. M. Augeri. 2013. "The Importance of Correcting for Sampling Bias in MaxEnt Species Distribution Models." *Diversity and Distributions* 19: 1366–1379. doi:10.1111/ddi.12096.

Lazo-Cancino, D., R. Rivera, K. Paulsen-Cortez, N. González-Berríos, R. Rodríguez-Gutiérrez, and E. Rodríguez-Serrano. 2020. "The Impacts Of Climate Change On The Habitat Distribution Of The Vulnerable Patagonian-Fueguian Species *Ctenomys Magellanicus* (Rodentia, Ctenomyidae)." *Journal of Arid Environments* 173: 104016. doi:10.1016/j.jaridenv.2019.104016.

Lazzaro, L., R. Bolpagni, G. Buffa, R. Gentili, M. Lonati, A. Stinca, A. T. R. Acosta, et al. 2020. "Impact Of Invasive Alien Plants On Native Plant Communities And Natura 2000 Habitats: State Of The Art, Gap Analysis And Perspectives In Italy." *Journal of Environmental Management* 274: 111140. doi:10.1016/j.jenvman.2020.111140.

Le Maitre, D. C., D. B. Versfeld, and R. A. Chapman, 2000. "Impact Of Invading Alien Plants On Surface Water Resources In South Africa: A Preliminary Assessment".

Liaw, A. and Wiener, M. 2002. "Classification and regression by random Forest." *R news* 2(3): 18–22.

Makaya, N. P., Mutanga, O., Kiala, Z., Dube, T. and Seutloali, K. E. 2019. "Assessing the potential of Sentinel-2 MSI sensor in detecting and mapping the spatial distribution of gullies in a communal grazing landscape." *Physics and Chemistry of the Earth, Parts A/B/C*, 112: 66–74. doi:10.1016/j.pce.2019.02.001

Makori, D. M., A. T. Fombong, E. M. Abdel-Rahman, K. Nkoba, J. Ongus, J. Irungu, G. Mosomtai, S. Makau, O. Mutanga, and J. Odindi. 2017. "Predicting Spatial Distribution Of Key Honeybee Pests In Kenya Using Remotely Sensed And Bioclimatic Variables: Key Honeybee Pests Distribution Models." *ISPRS International Journal of Geo-Information* 6: 66. doi:10.3390/ijgi6030066.

Malahlela, O. E., C. Adjorlolo, and J. M. Olwoch. 2019. "Mapping The Spatial Distribution Of *Lippia Javanica* (Burm. F.) Spreng Using Sentinel-2 And SRTM-Derived Topographic Data In Malaria Endemic Environment." *Ecological Modelling* 392: 147–158. doi:10.1016/j.ecolmodel.2018.11.020.

Masters, G., and L. Norgrove 2010. "Climate Change And Invasive Alien Species". *UK: CABI Working Paper*, 1.

Merow, C., Smith, M. J. and Silander Jr, J. A. 2013. "A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter." *Ecography* 36 (10): 1058–1069. doi:10.1111/j.1600-0587.2013.07872.x

Mkunyana, Y. P., D. Mazvimavi, S. Dzikiti, and Z. Ntshidi. 2018. "A Comparative Assessment Of Water Use By *Acacia Longifolia* Invasions Occurring On Hillslopes And Riparian Zones In The Cape Agulhas Region Of South Africa." *Physics and Chemistry of the Earth, Parts A/B/C* 112: 255–264. doi:10.1016/j.pce.2018.10.002.

Mohammadi, S., Ebrahimi, E., Moghadam, M. S. and Bosso, L. 2019. "Modelling current and future potential distributions of two desert jerboas under climate change in Iran." *Ecological Informatics* 52: 7–13. doi:10.1016/j.ecoinf.2019.04.003

Mokarram, M., G. Roshan, and S. Negahban. 2015. "Landform Classification Using Topography Position Index (Case Study: Salt Dome Of Korsia-Darab Plain, Iran)." *Modeling Earth Systems and Environment* 1: 1. doi:10.1007/s40808-015-0041-2.

Mooney, H. A. 2005. *Invasive Alien Species: A New Synthesis*. Island press.

Mtengwana, B., T. Dube, Y. P. Mkunyana, and D. Mazvimavi. 2020. "Use Of Multispectral Satellite Datasets To Improve

Ecological Understanding Of The Distribution Of Invasive Alien Plants In A Water-Limited Catchment, South Africa." *African Journal of Ecology* 58: 709–718. doi:10.1111/aje.12751.

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

Mudereri, B. T., E. M. Abdel-Rahman, T. Dube, T. Landmann, Z. Khan, E. Kimathi, R. Owino, and S. Niassy. 2020a. "Multi-Source Spatial Data-Based Invasion Risk Modeling Of Striga (*Striga Asiatica*) In Zimbabwe." *GIScience & Remote Sensing* 57(4): 553–571. doi:10.1080/15481603.2020.1744250

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

Naimi, B. (2020). "Package "Usdm". Accessed 18 February 2021]. https://cran.rproject.org/web/packages/sdm/sdm.pdf

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

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

Nath, A., Sinha, A., Lahkar, B.P. and Brahma, N. 2019. "In search of Aliens: Factors influencing the distribution of Chromolaena odorata L. and Mikania micrantha Kunth in the Terai grasslands of Manas National Park, India." *Ecological engineering* 131:16–26. doi:10.1016/j.ecoleng.2019.02.012

Ncube, B., M. D. Shekede, I. Gwitira, and T. Dube. 2020. "Spatial Modelling The Effects Of Climate Change On The Distribution Of *Lantana Camara* In Southern Zimbabwe." *Applied Geography* 117: 102172. doi:10.1016/j.apgeog.2020.102172.

Ndlovu, P., O. Mutanga, M. Sibanda, J. Odindi, and I. Rushworth. 2018. "Modelling Potential Distribution Of Bramble (*Rubus Cuneifolius*) Using Topographic, Bioclimatic And Remotely Sensed Data In The Kwazulu-Natal Drakensberg, South Africa." *Applied Geography* 99: 54–62. doi:10.1016/j.apgeog.2018.07.025.

Ng, W. T., A. Candido De Oliveira Silva, P. Rima, C. Atzberger, and M. Immitzer. 2018. "Ensemble Approach For Potential Habitat Mapping Of Invasive *Prosopis* Spp. In Turkana, Kenya." *Ecology and Evolution* 8: 11921–11931. doi:10.1002/ece3.4649.

O'donnell, M. S., and D. A. Ignizio. 2012. "Bioclimatic Predictors For Supporting Ecological Applications In The Conterminous United States." *U.S. Geological Survey Data Series* 691(10): 4–9.

Orimoloye, I. R., O. O. Ololade, S. P. Mazinyo, A. M. Kalumba, O. Y. Ekundayo, E. T. Busayo, A. A. Akinsanola, and W. Nel. 2019. "Spatial Assessment Of Drought Severity In Cape Town Area, South Africa." *Heliyon* 5: e02148. doi:10.1016/j.heliyon.2019.e02148.

Otieno, B. A., H. F. Nahrung, and M. J. Steinbauer. 2019. "Where Did You Come From? Where Did You Go? Investigating The Origin Of Invasive *Leptocybe* Species Using Distribution Modelling." *Forests* 10: 115. doi:10.3390/f10020115.

Pang, S. E. H., J. D. T. De Alban, and E. L. Webb. 2021. "Effects of Climate Change and Land Cover on the Distributions of a Critical Tree Family in the Philippines." *Scientific Reports* 11: 276. doi:10.1038/s41598-020-79491-9.

Pearce, J. L., and M. S. Boyce. 2006. "Modelling Distribution And Abundance With Presence-Only Data." *Journal of Applied Ecology* 43: 405–412. doi:10.1111/j.1365-2664.2005.01112.x.

Perring, F. 1956. "Climatic And Edaphic Gradients Of Chalk Grassland." PhD, Queens College, Cambridge Unversity.

Perring, F. 1959. "Topographical Gradients Of Chalk Grassland." *Journal of Ecology* 47: 447–481.

Phillips, S. J., Anderson, R. P. and Schapire, R. E. 2006. "Maximum entropy modeling of species geographic distributions." *Ecological modelling* 190(3–4): 231–259. doi:10.1016/j.ecolmodel.2005.03.026

Prinsloo, F. W., and D. F. Scott. 1999. "Streamflow Responses To The Clearing Of Alien Invasive Trees From Riparian Zones At Three Sites In The Western Cape Province." *The Southern African Forestry Journal* 185: 1–7. doi:10.1080/10295925.1999.9631220.

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

Richard, K., Abdel-Rahman, E. M., Mohamed, S. A., Ekesi, S., Borgemeister, C. and Landmann, T. 2018. "Importance of remotely-sensed vegetation variables for predicting the spatial distribution of African citrus triozid (Trioza erytreae) in Kenya." *ISPRS International Journal of Geo-Information* 7(11): 429. doi:10.3390/ijgi7110429

R Core Team. 2019. "R: A language and environment for statistical computing." *R Foundation for Statistical Computing, Vienna, Austria*. https://www.R-project.org/ .

Rouget, M., D. M. Richardson, J. L. Nel, D. C. Le Maitre, B. Egoh, and T. Mgidi. 2004. "Mapping The Potential Ranges Of Major Plant Invaders In South Africa, Lesotho And Swaziland Using Climatic Suitability." *Diversity and Distributions* 10: 475–484. doi:10.1111/j.1366-9516.2004.00118.x.

Seif, A. 2014. "Using Topography Position Index For Landform Classification (Case Study: Grain Mountain)." *Bulletin of Environment, Pharmacology and Life Sciences* 3: 2277.

Serdeczny, O., S. Adams, F. Baarsch, D. Coumou, A. Robinson, W. Hare, M. Schaeffer, M. Perrette, and J. Reinhardt. 2016. "Climate Change Impacts In Sub-Saharan Africa: From Physical Changes To Their Social Repercussions." *Regional Environmental Change* 17: 1585–1600. doi:10.1007/s10113-015-0910-2.

Sher, A. A., K. Wiegand, and D. Ward. 2010. "Do *Acacia* And *Tamarix* Trees Compete For Water In The Negev Desert?" *Journal of Arid Environments* 74: 338–343. doi:10.1016/j.jaridenv.2009.09.007.

Somodi, I., N. Lepesi, and Z. Botta-Dukát. 2017. "Prevalence Dependence In Model Goodness Measures With Special

Emphasis On True Skill Statistics." *Ecology and Evolution* 7: 863–872. doi:10.1002/ece3.2654.

Sörensen, R., Zinko, U. and Seibert, J. 2006. "On the calculation of the topographic wetness index: evaluation of different methods based on field observations." *Hydrology and Earth System Sciences* 10(1): 101–112.

Stohlgren, T. J., P. Ma, S. Kumar, M. Rocca, J. T. Morisette, C. S. Jarnevich, and N. Benson. 2010. "Ensemble Habitat Mapping Of Invasive Plant Species." *Risk Analysis* 30: 224–235. doi:10.1111/j.1539-6924.2009.01343.x.

Tarabon, S., R. Bertrand, C. Lavoie, T. Vigouroux, F. Isselin-Nondedeu, and G. Fried. 2018. "The Effects of Climate Warming and Urbanised Areas on the Future Distribution of *Cortaderia Selloana*, Pampas Grass, in France." *Weed Research* 58: 413–423. doi:10.1111/wre.12330.

Terzano, D., I. Kotzé, C. Marais, S. Cianciullo, A. Farcomeni, P. Caroli, L. Malatesta, and F. Attorre. 2018. "Environmental And Anthropogenic Determinants Of The Spread Of Alien Plant Species: Insights From South Africa's Quaternary Catchments." *Plant Ecology* 219: 277–297. doi:10.1007/s11258-018-0795-5.

Thalmann, D. J. K., D. Kikodze, M. Khutsishvili, D. Kharazishvili, A. Guisan, O. Broennimann, and H. Müller-Schärer. 2015. "Areas Of High Conservation Value In Georgia: Present And Future Threats By Invasive Alien Plants." *Biological Invasions* 17: 1041–1054. doi:10.1007/s10530-014-0774-2.

Truong, T. T., G. E. S. J. Hardy, and M. E. Andrew. 2017. "Contemporary Remotely Sensed Data Products Refine Invasive Plants Risk Mapping In Data Poor Regions." *Frontiers in Plant Science* 8: 770. doi:10.3389/fpls.2017.00770.

Vorsino, A. E., L. B. Fortini, F. A. Amidon, S. E. Miller, J. D. Jacobi, J. P. Price, S. O. O. A. Gon Iii, and G. A. Koob. 2014. "Modeling Hawaiian Ecosystem Degradation Due To Invasive Plants Under Current And Future Climates." *PLoS ONE* 9(5): e95427. doi: 10.1371/journal.pone.0095427

Warren, D. L., N. J. Matzke, and T. L. Iglesias. 2019. "Evaluating Presence-Only Species Distribution Models With Discrimination Accuracy Is Uninformative For Many Applications." *Journal Of Biogeography* 47(1): 167–180. doi:10.1101/684399.

Weiss, A. 2001. "Topographic Position And Landforms Analysis". Poster Presentation, Esri User Conference, San Diego, CA.

West, A. M., P. H. Evangelista, C. S. Jarnevich, S. Kumar, A. Swallow, M. W. Luizza, and S. M. Chignell. 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." *International Journal of Applied Earth Observation and Geoinformation* 59: 135–146. doi:10.1016/j.jag.2017.03.009.

Wilcove, D. S., D. Rothstein, J. Dubow, A. Phillips, and E. Losos. 1998. "Quantifying Threats To Imperiled Species In The United States." *BioScience* 48: 607–615. doi:10.2307/1313420.

Ye, X., X. Yu, C. Yu, A. Tayibazhaer, F. Xu, A. K. Skidmore, and T. Wang. 2018. "Impacts of Future Climate and Land Cover Changes on Threatened Mammals in the Semi-arid Chinese Altai Mountains." *Science of the Total Environment* 612: 775–787. doi:10.1016/j.scitotenv.2017.08.191.

Yu, H., A. R. Cooper, and D. M. Infante. 2020. "Improving Species Distribution Model Predictive Accuracy Using Species Abundance: Application With Boosted Regression Trees." *Ecological Modelling* 432: 109202. doi:10.1016/j.ecolmodel.2020.109202.

Zhang, L., F. Huettmann, S. Liu, P. Sun, Z. Yu, X. Zhang, and C. Mi. 2019. "Classification And Regression With Random Forests As A Standard Method For Presence-Only Data Sdms: A Future Conservation Example Using China Tree Species." *Ecological Informatics* 52: 46–56. doi:10.1016/j.ecoinf.2019.05.003.