Physics and Chemistry of the Earth 135 (2024) 103633

Contents lists available at ScienceDirect

Physics and Chemistry of the Earth

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

Modelling lantana camara invasion in the inkomati catchment in Mpumalanga, South Africa

Vuyelwa Emmaculate Mtyobila *, Cletah Shoko

Division of Geography, School of Geography, Archaeology and Environmental Studies, University of the Witwatersrand, 1 Jan Smuts Avenue, Braamfontein, 2000, Johannesburg, South Africa

ARTICLE INFO

Keywords:
Lantana Camara
MaxEnt
Satellite data
Species distribution models
Topographic variable

ABSTRACT

Lantana Camara ranks among the most notorious and hazardous invasive species on the planet. There has been extensive research focusing on the remote sensing of *Lantana Camara* (*L. Camara*). However, the factors that facilitate its spread, especially in savanna rangelands, are not yet fully comprehended. The present research investigated the recent spatial distribution of *L. Camara* in the Inkomati catchment, Mpumalanga, South Africa. The research used the Random Forest classification algorithm, with Sentinel-2 remotely sensed data within Google Earth Engine. The study also modelled areas vulnerable to *L. Camara* invasion using the Maxent species distribution model. The study found that *L. Camara* covered approximately 34.86% of the entire study area, with a user's and a producer's accuracies of 91% and 84%, respectively. Additionally, elevation was identified as the most influential topographic variable on the species spatial distribution and invasion, while Topographic Wetness Index had the least influence. The model developed using topographic variables had the highest accuracy of 0.88. On the other hand, the predictive model that utilised Sentinel 2 bands had the least accuracy of 0.81. The red edge and NIR bands with Random Forest allowed for an accurate assessment of *L. Camara* distribution. The study therefore provides the basis for the control or further expansion of *L. Camara* species in the province.

1. Introduction

Invasive Alien Species (IAPs) are weeds relocated from their natural environment to a new one (Rai and Singh, 2020). When established, IAPs have the capability to swiftly spread across large areas (Ricciardi and Cohen, 2007). These species continue to be a cause for concern for biological conservationists, ecologists, and managers of natural resources because of their fast expansion, risk to biodiversity, and natural ecosystems. Invasive species alter the nutrient cycle, hydrology, and carbon sequestration processes (Pimentel et al., 2005; Polley et al., 1997) homogenize natural ecosystems (Mooney and Hobbs, 2000), and are recognized as a major contributor to the loss of biodiversity worldwide (Wilcove and Chen, 1998). These species also cause severe economic losses associated with their eradication or control (Pimentel et al., 2005). For instance, Australia on its own loses roughly 2.2 million USD each year (Cusack et al., 2009) whereas the United States loses an estimated 120 billion USD per year (Ayele, 2007). The economic costs associated with livestock poisoning by *Lantana Camara* (*L. Camara*) in South Africa are projected to be R 1 728 900 each year (Heystek, 2006).

Historically, aerial images and field surveys have been employed to obtain information about IAPs (Zuberi et al., 2014). However, several studies have emphasized that such techniques are not sustainable since they are time consuming, expensive and labour-intensive, particularly for large-scale applications (Shariq and Hughes, 2020). Conversely, remote sensing has grown in recognition as a feasible approach in IAPs mapping (Gómez- Casero et al., 2010). Various multispectral (Peerbhay et al., 2016; Adam et al., 2017; Dhau, 2008; Oumar, 2016), hyperspectral (Peerbhay et al., 2016; Underwood et al., 2003), unmanned aerial vehicles (Iqbal et al., 2023; Niphadkar, 2016), and aerial images (Evritt et al., 2018; Tharushi et al., 2018; Sundaram et al., 2012), have been successfully used to identify invasive species. However, high spatial resolution datasets, like aerial images and unmanned aerial vehicles (UAVs), as well as commercially acquired hyperspectral and multispectral datasets are limited by their ability to capture only a small area at a time, and high acquisition cost, making it difficult to map invasive species across large areas (Kattenborn et al., 2020). In addition, intermediate spatial resolution sensors like the Landsat and SPOT-5 are limited by their inability to detect small and isolated patches (Zhang and

* Corresponding author.
E-mail address: vuyelwa.mtyobila1@gmail.com (V.E. Mtyobila).

https://doi.org/10.1016/j.pce.2024.103633
Received 16 September 2023; Received in revised form 28 November 2023; Accepted 12 May 2024
Available online 17 May 2024
1474-7065/

Foody, 1998).

Recent improvements in satellite remote sensing and cloudy computing technology have brought about new imaging options for invasive species detection. For example, Google Earth Engine cloud computing platform enables the user to have access to different remote sensing data and machine learning algorithms for advanced monitoring of vegetation invasion. Similarly, the spectral, temporal and spatial properties of Sentinel 2 make it suitable for accurate mapping of invasive species, as demonstrated by Rima et al. (2017). On the other hand, Spatial Distribution Models (SDMs) are becoming prominent in ecological remote sensing for identifying areas vulnerable to invasion (Martins et al., 2016). For example, Huiyu et al. (2020) successfully investigated the influence of global climatic suitability on the distribution of *Spartina alterniflora* using species distribution models (SDMs). Despite its effectiveness in detecting and mapping invasive species, there has been limited research on the use of Sentinel-2 to identify and map *L. Camara* using SDMs. Therefore, this study aimed to explore *L. Camara* spatial patterns in the Inkomati catchment, in Mpumalanga, South Africa. The study had three specific objectives, which were to: (i) map the recent spatial distribution of *L. Camara*; (ii) establish key topographic variables and their influence on *L. Camar*a spatial distribution and invasion and (iii) model areas vulnerable to its invasion in the Inkomati catchment.

2. Methodological approach

2.1. Study area

The Inkomati catchment is located in the Mpumalanga Province, in South Africa (as shown in Fig. 1). The IWMA spans an area of 28 757 km2 and comprises the Crocodile, Komati, and the Sabie major rivers (Mahlathi, 2018). The climate in the study area is highly diverse and influenced by the terrain, ranging from temperate Highveld in the west to sub-tropical in the eastern Lowveld (McLoughlin et al., 2021). The annual average temperature is 17 °C, with January being the hottest month with an average of 21 °C. June is the coldest month characterised by minimum average temperature of 11.5 °C (Hassan and Mahlathi, 2020). Frost is usually observed from June to early August, except in the extreme east where there is no frost. The region experiences most of its rainfall from October to April, with an average annual precipitation of 767 mm. There has been reported reduction of natural drainage, primarily due to large commercial plantations and invasive alien plants covering approximately 132 000 ha (Raheem, 2013). Afforestation and invasive alien vegetation are estimated to have reduced runoff by considerable amount. According to DWAF (2004), alien vegetation absorbs 91 million m3/year of water in the IWMA, with 7 million m3/year in Komati West, 0 million m3/year in Lower Komati, 57 million m3/year in Crocodile, and 24 million m3/year in Sabie, and 3 million m3/year in Sand.

2.2. Field data collection

In this study, previously gathered field data on *L. Camara*, land use and land cover types were utilized. To accurately record the location points of *L. Camara*, a handheld Garmin eTrex Global Positioning System (GPS) receiver was utilized with sub-meter precision in mid- December 2020. The study consisted of five classes, namely waterbody, *L. Camara*, cashnuts, banana, and sugarcane plantations, with 284, 225, 80, 170, and 122 samples for each class, respectively. 70% of the sampling points were used for training, while 30% were used for model validation, following the approach used in previous studies (Maxwell et al., 2018). GPS coordinates were captured using Microsoft Excel Version 4.0 and presented in a table format. Then, they were overlaid on the study area shapefile in the ArcGIS 10.6 software environment.

2.3. Sentinel-2 satellite imagery acquisition

Sentinel-2 imagery, which is freely available and consists of the two satellites Sentinel-2A and Sentinel-2B, is a multispectral sensor that was launched on June 23, 2015 (Paul and Kumar, 2019). An aggressive invasive species like *L. Camara* can now be detected, mapped, and monitored thanks to this sensor. Sentinel-2 data demonstrate the necessity and viability of employing broadband multispectral sensors to characterize and profile invasive species, a previously challenging task. The satellite acquires data every five days, covering 13 spectral bands (443–2190 nm) with various spatial resolutions and a swath width of 290 km. Table 1 highlights Sentinel 2 data characteristics. The study did not include bands 1, 9, or 10 because of their 60-m- wide coarse spatial resolution. The image was atmospherically corrected using the Sen2Cor tool in the Sentinel Application Platform (SNAP) and spatially

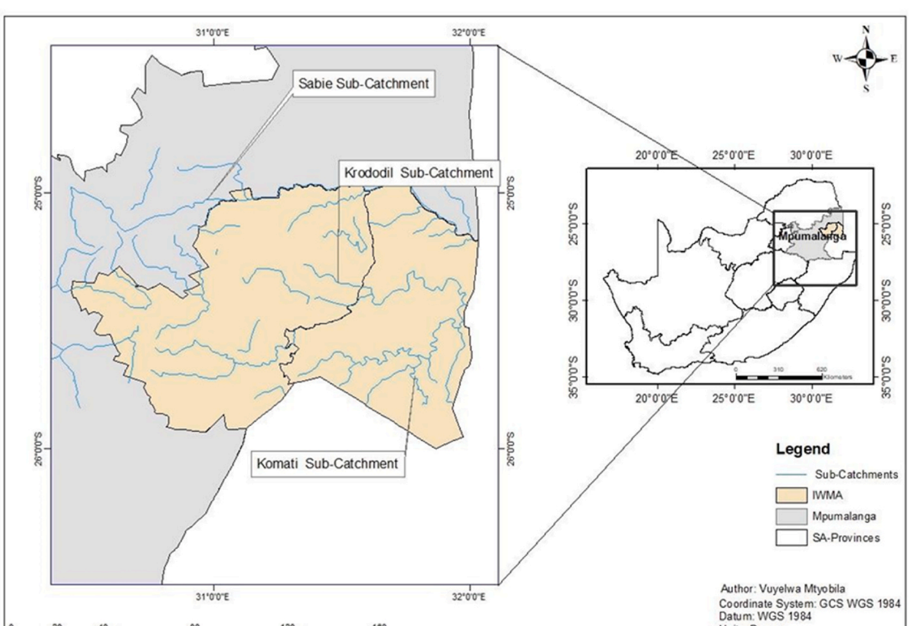

Fig. 1. Map of the study area in Inkomati Catchment in Mpumalanga, South Africa.

Table 1
Spectral characteristics of Sentinel-2 datasets.

Sentinel-2 Bands	Central Wavelength (μm)	Resolution (m)
Band 1 – Coastal aerosol	0.443	60
Band 2 – Blue	0.490	10
Band 3 – Green	0.560	10
Band 4 – Red	0.665	10
Band 5 – Vegetation Red Edge	0.705	20
Band 6 – Vegetation Red Edge	0.740	20
Band 7 – Vegetation Red Edge	0.783	20
Band 8 – NIR	0.842	10
Band 8 A – Vegetation Red Edge	0.865	20
Band 9 – Water Vapour	0.945	60
Band 10 – SWIR – Cirrus	1.375	60
Band 11 – SWIR	1.610	20
Band 12 - SWIR	2.190	20

resampled to 10 m in Google Earth Engine (GEE).

2.4. Topographic data acquisition and preparation

The Advanced Space-born Thermal Emission and Reflection Radiometer (ASTER), (https://asterweb.jpl.nasa.gov/gdem.asp), digital elevation model (DEM) with a 30 m spatial resolution was acquired to derive topographic variables. Elevation, Topographic Position Index (TPI), and the Topographic Wetness Index (TWI) were all extracted from the DEM using the spatial analyst tool in ArcGIS. TWI, a soil moisture index, has been used successfully in earlier studies to analyze vegetation patterns and forecast their distribution (Sorensen and Seibert, 2007). It affects the growth and composition of vegetation. TWI indicates the likelihood of a wet area higher values indicate more moisture/water content compared to lower values (Davidson et al., 1998). TIP is also a useful tool for classifying the landforms within a specific location and for understanding the biophysical processes. This information is important when predicting the suitability of habitats and the distribution of species in the area (Seif, 2014). A value close to zero indicates a flat slope, while a large positive value indicates that the central pixel is located on a hill or ridge. Conversely, a large negative value suggests a valley or gulley (Weiss, 2001). Slope and aspect were derived from the DEM as well. Soil moisture, soil type, precipitation, and sun angle are influenced by terrain variables, hence the distribution of vegetation. These topographic variables have an impact on hydrological processes and solar radiation, causing changes in soil moisture, plant function, as well as structure (Guo et al., 2020).

2.5. Image classification for the recent spatial distribution of Lantana Camara

The RF model was implemented in Google Earth Engine with a Sentinel-2 image to highlight the spatial coverage of *L. Camara* and other land use and land cover. This technique is frequently applied in image classification to increase precision and decrease overfitting (Odindi et al., 2016; Lee et al., 2014). Out-of-bag (OOB) samples are samples that are not part of the bootstrap sample (Lee et al., 2015; Adelabu et al., 2013). The accuracy of the classification model is evaluated using these OOB samples, which make up one-third of the dataset.

2.6. Accuracy assessment of digital image classification

To evaluate the accuracy of the image classification, Sentinel-2 MSI data was employed, and confusion matrices were used to calculate the User Accuracy (UA), Overall Accuracy (OA), and Producer Accuracy (PA). 70% of the training data was used to generate the *L. Camara* maps, while the remaining 30% was utilized to evaluate classification accuracy.

2.7. Modelling areas vulnerable to the Lantana Camara invasion

Vulnerable areas to the invasion of L. *Camara* were modelled using the Maximum Entropy (MaxEnt) model freely accessible online (http://biodiversityinformatics.amnh.org/open_source/MaxEnt/). MaxEnt is a stand-alone Java program created for modelling species distributions using occurrence records of the species and environmental variables (Phillips and Dudík, 2008; Phillips et al., 2006). MaxEnt is an SDM that is helpful in locating the distribution of invasive species and has been acknowledged as one of the top SDMs in invasion modelling (Ficetola et al., 2007). The model is sensitive to spatial errors associated with low data and can also handle different datasets (Phillips et al., 2006). MaxEnt has proved successful in modelling and forecasting the distribution of invasive species (e.g., Choudhury et al., 2016; Qin et al., 2016; Roger et al., 2015). The field dataset was divided into two portions: 20% for model testing and 80% for model development. A jack-knife test was also used to assess the influence of the predictor variables in the species distribution and their distinctive contributions. Sentinel-2 bands, elevation, slope, TPI, and TWI were the variables used in the MaxEnt model to model the distribution of the species.

2.8. Evaluation of MaxEnt's accuracy in modelling areas vulnerable to invasion

The accuracy and effectiveness of the MaxEnt model were determined using the Area Under Curve (AUC) and True Skill Static (TSS). The AUC evaluates the congruence between the species' observed presence and the classifier's predicted distribution, demonstrating whether the relative importance of presence versus absence was correctly arranged. Low AUC values from 0.5 to 0.7 indicates random to poor performance, whereas from 0.7 to 0.9 illustrates moderate performance, and 0.9 indicates high performance (Shabani et al., 2016). TSS values range from −1 to 1, with values closer to 1 indicating better model performance and those closer to −1 indicating poor and random model performance (Mouton et al., 2010). Further, with background samples as absence data, the error matrix was used to calculate the specificity, Kappa, sensitivity, and TSS values. The 10th percentile was used to gauge the precision of the classification.

3. Results

This investigated the recent spatial distribution of *L. Camara* and predicted areas which are vulnerable to its invasion. The study used Sentinel-2 satellite image, Random Forest classification algorithm in Google Earth Engine, to understand the recent spatial distribution as well as the MaxEnt model to model vulnerability of areas to invasion.

3.1. The recent spatial coverage of Lantana Camara

Fig. 2 illustrates the recent spatial coverage of *L. Camara* and other land use land cover types. Overall, *L. Camara* is scattered throughout the study area. However, the majority of *L. Camara* is in the northeast and central parts of the area. Additionally, *L. Camara* was found to be in closer proximity to waterbodies, as well as sugarcane and banana plantations.

The percentage areal coverage of each class is presented in Fig. 3. *L. Camara* had the highest coverage of 34.86%, followed by sugarcane plantation at 21.44%. In contrast, banana plantation had the least coverage at 11.64%.

3.2. Sentinel-2 bands variable importance in image classification

Fig. 4 presents random forest bands variable importance in classifying *L. Camara* and other classes. Band 8a (Vegetation Red Edge) and Band 11 (Shortwave Infrared) had the highest importance, respectively. In contrast, Band 4 (Red) had the lowest importance score of 340,

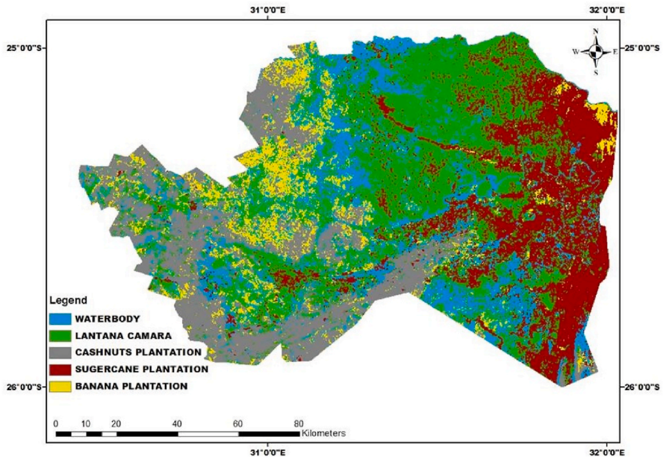

Fig. 2. Distribution of L. Camara and other land cover types across the study area.

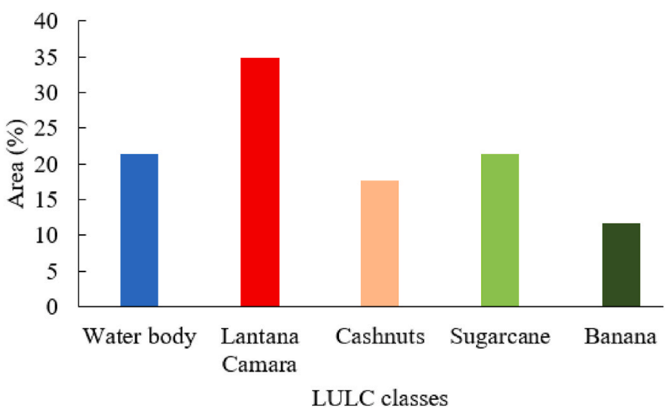

Fig. 3. Areal coverage of *L. Camara* and other LULC classes within the study area.

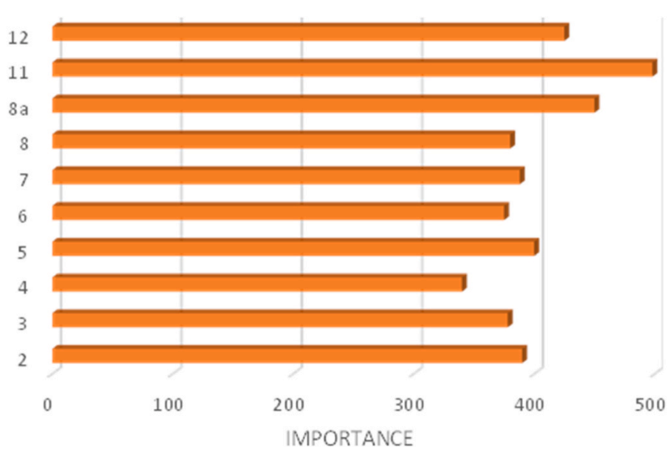

Fig. 4. Random forest bands variable importance.

indicating its lesser significance in the classification.

3.3. Random forest classification algorithm accuracy assessment results

Table 2 displays the confusion matrix that illustrates the classification of *L. Camara* and other land cover types using the Random Forest classification algorithm. The classification was associated with an overall accuracy of 84% and a kappa of 0.78. The UA for *L. Camara* was 91%, while the PA was 84%.

3.4. Modelling areas vulnerable to invasion

Fig. 5 (a) illustrates vulnerability map of *L. Camara* obtained using elevation, slope, TPI, and TWI. The areas with high invasion risks are mostly located on the west, while the areas with low invasion risks are situated on the east. Fig. 5 (b) presents a vulnerability map of *L. Camara* using Sentinel-2 bands. The invasion risks are moderate in almost all parts of the map except for the southwestern and north-western edges. Both vulnerability maps identify the southern edge of the study are to be less vulnerable to *L. Camara* invasion.

3.5. MaxEnt's accuracy in modelling areas vulnerable to L. Camara invasion

Fig. 6 illustrates variable importance in predicting areas which have been identified as vulnerable to invasion by *L. Camara*. The use of topographic variables in isolation achieved the highest predictive accuracy, with an AUC of 0.882, whereas the use of Sentinel-2 had a lower accuracy, with an AUC of 0.81. Elevation was identified as the most important variable in predicting vulnerable areas. Conversely, the TWI was the least influential topographic variable. Fig. 6 (b) revealed that the Vegetation Red Edge bands (6 and 7) and the Near Infrared band (band 8) were identified as most important variables, while the red band (4) was identified as the least important variable in predicting vulnerable areas.

4. Discussion

The aim of this study was to map the recent distribution of *L. Camara* in the Inkomati catchment in Mpumalanga, South Africa, using the Random Forest classification algorithm and Sentinel-2 data. The study also predicted areas vulnerable to *L. Camara* invasion using topographical variables and MaxEnt species distribution model, and identified the most influential variable in prediction of these areas.

4.1. Mapping the recent spatial distribution of L. Camara in the inkomati catchment

The majority of *L. Camara* was detected in the northeast region of the study area, although the species seemed to be scattered across the study area. This distribution can be explained by different factors which facilitate the growth of the species. *L. Camara* is known to thrive in warm, dry environments with plenty of sunlight. If one side of the study area has more of these environmental conditions compared to the other side, it could explain why the plant is distributed unevenly (Negi et al., 2019). *L. Camara* prefers well-drained soils with a slightly acidic pH. If the soil on one side of the study area meets these requirements better than the other side, it could explain why the plant is found there and not in other areas (Mandal and Joshi, 2014). *L. Camara* is often introduced to new areas through human activities, such as gardening or agriculture. If humans have been more active on one side of the study area compared to the other side, it could explain why the plant is distributed unevenly (Richardson and Rejmánek, 2004). *L. Camara* can compete with other plant species for resources such as water or nutrients. If other plants are more abundant on one side of the study area, it could limit the growth and distribution of *L. Camara* in that area (Mandal and Joshi, 2015). *L. Camara* can produce large amounts of seeds that are dispersed by animals, water, or wind. If the distribution of the plant is influenced by the movement of these dispersal agents, it could explain why the plant is found in certain areas and not in others (Kohli et al., 2006). Additionally, the topography of the study area could play a role in the plant's distribution. For example, areas with more slopes or changes in elevation may be less favourable for *L. Camara* growth (Thapa et al., 2018).

Table 2
Confusion Matrix associated with image classification.

Class Name	Waterbody	*L. Camara*	Cashnuts	Sugarcane	Banana	Total	UA
Waterbody	33	6	2	2	1	44	75
L. Camara	2	41	0	1	1	45	91
Cashnuts	0	0	3	0	1	4	75
Sugarcane	6	2	0	21	0	29	72
Banana	1	0	0	1	34	36	94
Total	42	49	5	25	37	158	
PA	78	84	60	84	92		
Overall Accuracy: 84%							
Kappa coefficient: 0.78							

*PA: Producer Accuracy, UA: User Accuracy.

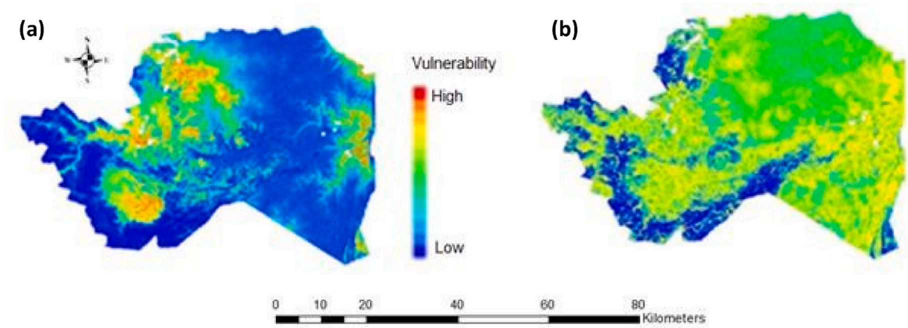

Fig. 5. Vulnerability maps derived from (a) topographic variables and (b)Sentinel-2 bands.

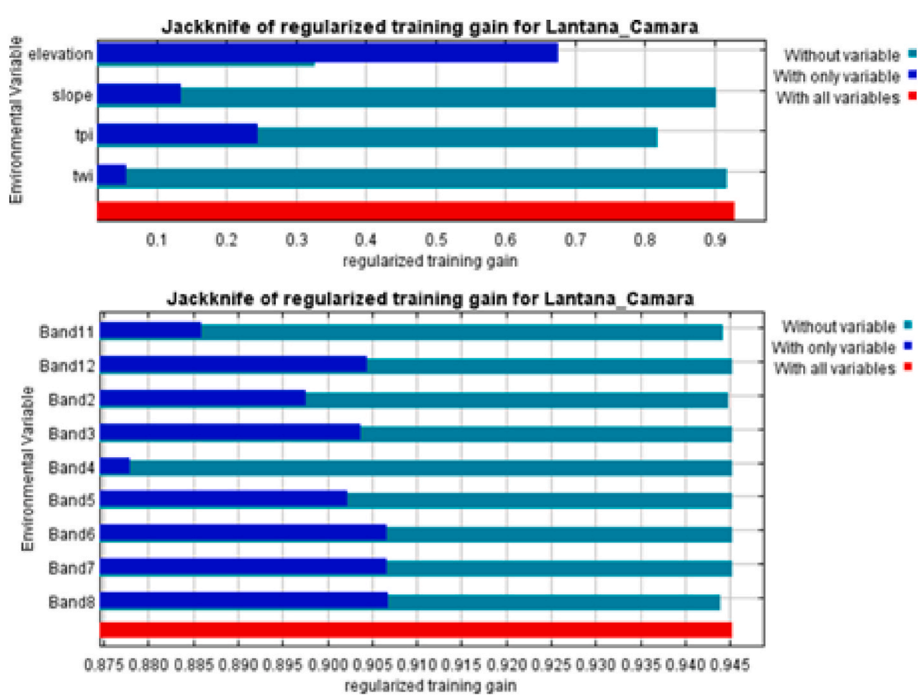

Fig. 6. Variable importance using: (a) topographic variables and (b) Sentinel 2 spectral ands.

Natural disturbances such as fires, floods, or landslides could create favourable conditions for *L. Camara* growth and distribution in certain areas. Conversely, areas that are less disturbed may be less favourable for the plant (Sukopp and Starfinger, 1999).

In addition, there have been several studies that have examined the distribution of *L. Camara*. Singh et al. (2008) found that *L. Camara* was more abundant in areas with warmer temperatures and higher levels of solar radiation, with lower annual rainfall. In a different study, DiTommaso et al. (2005) found that L. Camara mostly occur in areas with sandy soils, with low pH, nitrogen and phosphorus concentration. Rathour et al. (2017) found that human activity, such as deforestation and agriculture, was a key factor in the spread of *L. Camara* in the Western Ghats region of India. Ospina-Torres et al. (2019) found that *L. Camara* seeds were dispersed by several animal species, including birds and bats, and that seed dispersal was an important factor in the spread of the plant in Colombia. Overall, these studies suggest that the survival and growth of *L. Camara* at different areas is influenced by an ensemble of environmental factors, soil type, human activity, competition with other plants, and seed dispersal. However, it remains crucial to understand that the specific factors that influence the distribution of the plant vary by location and context of the study.

4.2. Establishing key topographic factors and their influence on L. Camara invasion

Elevation was identified as the most important topographic variable to model *L. Camara* distribution. Elevation influences the distribution of the species due to variations in temperature and soil conditions along an elevation gradient. Rainfall and temperature patterns are critical factors affecting the distribution of *L. Camara*, and elevation can influence these factors. Additionally, soil properties, such as nutrient content, moisture levels, and pH, can be affected by elevation, which, in turn, can influence the growth and survival of the species (Phillips and Dudík, 2008). Studies by Adeola (2017) and Ndlovu et al. (2018) have found that elevation is a significant factor in explaining the occurrence of *L. Camara*. Adeola (2017) further asserts that elevation facilitates the distribution of plant species and soil conditions. In another study, Priyanka and Joshi (2013) also supports the importance of elevation gradients in predicting the distribution of *L. Camara*, as the species is known to thrive at lower altitudes.

The TWI had the least importance on *L. Camara* spatial coverage. One possible explanation is that TWI may not be a significant variable in the areas where the species is found, as it is primarily used to measure the wetness of soil and may not be as relevant in drier regions. Additionally, other factors, such as soil nutrient content or temperature, may facilitate the survival of the species in these areas. Several studies have explored the relationship between TWI and *L. Camara* with mixed findings. In the Kumaon region of India, Singh et al. (2019) found significant influence of TWI on *L. Camara* spatial distribution. They observed that the species occurred more frequently on sites with higher TWI, which suggests that soil moisture plays a crucial role in the growth and survival of the species. In support of this, research by Vila et al. (2010) found that *L. Camara* is adaptable to a range of soil types and moisture levels, which may explain why TWI was found to have a weaker influence on its distribution compared to other topographic variables. Similarly, a study by Chauhan et al. (2010) found that temperature and precipitation patterns were significant drivers of the species' distribution, whereas soil moisture was not found to be as important.

Contrary to this study, other studies have found little to no relationship between TWI and *L. Camara* distribution. Krishnakumar et al. (2018) in India reported that TWI was not a significant predictor of the species' distribution. The authors observed that soil moisture content does not limit the occurrence of the species in the region. Similarly, a study conducted in the Tarai region of Nepal by Devkota et al. (2017) found that TWI was not a significant predictor of the distribution of *L. Camara*. The authors found that the species occurred across a wide range of TWI values, which suggests that soil moisture is not a limiting factor for the species in the region. The results of the research conducted by Devkota et al. (2017) and Krishnakumar et al. (2018) concur with the findings of the present study. Overall, the relationship between TWI and *L. Camara* distribution vary depending on the region and local environmental conditions. While some studies have found a significant relationship between TWI and the species' distribution, others have found little to no relationship. Further research is needed to better explore the role of TWI in the growth and survival of *L. Camara*, possibly across seasonal variations.

4.3. Modelling areas vulnerable to L. Camara invasion in the inkomati catchment

The findings indicate that the central-western section of the study area is more vulnerable to *L. Camara* invasion than other parts. Areas which were associated with low elevation, a gentle slope, low TPI and low TWI within the area, had high invasion risks, while areas that are more elevated, have a steep slope, a high TPI and high TWI have low invasion taking place. This may be related to the influence of elevation on the spatial distribution of *L Camara*. Temperature, precipitation, and soil properties are all influenced by elevation, which can affect the distribution and survival of the species (Munson et al., 2011). On the other hand, the higher the elevation, the cooler the temperature tends to be, thereby hindering vegetation growth (Ab Lah et al., 2021). Additionally, higher elevations may experience more snowfall, which can affect soil moisture levels and contribute to the formation of glaciers and permafrost. Soil properties can also be influenced by elevation. Higher elevations generally have thinner soils due to the presence of rocks and steeper slopes that prevent the accumulation of organic matter (Yang et al., 2018). Additionally, soil pH may be lower in high-elevation areas due to increased precipitation, which can leach minerals from the soil. However, there may be variations in soil properties depending on the local geology and climate, so it is important to consider specific factors when studying the relationship between elevation and soil (Buytaert et al., 2011). In a study by Rodgers and Parker (2003), the level of alien plant invasion was compared between islands in the United States. It was found that locations with high levels of damage had significantly greater alien plant cover in both ecosystems and islands. Another study by Lin (2005) focused on *L. Camara* cover on main roadsides in Moorea, French Polynesia, in relation to environmental factors. The study found that *L. Camara* covered 1.99% of the roadside area, with the greatest presence being in areas with agricultural disturbance and associated with the roadside habitat type. Additionally, Muniappan et al. (2004) found that *L. Camara* was more abundant in disturbed habitats at lower elevations in the Pacific Island of Guam. Bhowmik et al. (2015) found that *L. Camara* occurred in areas with high levels of disturbance and lower elevations in the Chittagong Hill Tracts of Bangladesh. Chandra et al., 2006 found that *L. Camara* was more abundant in disturbed habitats such as roadsides, wastelands, and forest edges at lower elevations in the Western Himalayan region of India. Ntuli et al. (2007) found that *L. Camara* was more abundant in disturbed habitats such as roadsides and clearings at lower elevations in the Makhonjwas Mountains of South Africa. Several studies have found that L. Camara tends to occur at lower elevations, where disturbance is more common (Rathour et al., 2016; Rasool et al., 2017)

In addition, the present study area is characterized by a range of habitats, including grasslands, savannas, forests, and wetlands, and is home to a variety of species. Based on the studies discussed earlier, *L. Camara* is more likely to occur in disturbed habitats such as roadsides, clearings, and forest edges, which are more common at lower elevations in the study area. Therefore, areas that have experienced high levels of human disturbance, such as agricultural lands, urban areas, and transportation corridors, may be at a higher risk of invasion by *L. Camara*. Additionally, at lower elevations, higher temperatures and lower rainfall, may also provide more favourable conditions for the growth and spread of *L. Camara* in the study area. These factors require attention when assessing the potential risk of invasion by *L. Camara* in the region and when developing strategies for managing and controlling this invasive species.

The study utilized two sets of variables and models to predict areas vulnerable to *L. Camara* invasion. Model 1, which employed topographic variables, produced the most accurate predictions with an AUC of 0.882. Model 2, which used Sentinel-2 bands, yielded lower accuracies with an AUC of 0.810. Previous studies have found that models based on topographic variables perform better than those based solely on Sentinel-2 variables (Dube et al., 2022; Malahlela et al., 2022; Ndlovu et al., 2022; Parra et al., 2004). For instance, Dube et al. (2022) used MaxEnt, Sentinel-2 derived variables and environmental variables to model the distribution of *L. Camara*, and the model incorporating environmental variables achieved the highest accuracy. Similarly, Parra et al. (2004) utilized climatic, topographic (elevation), and Normalized Difference Vegetation Index to predict the distribution of species in the Ecuadorian

Andes, South America. Models including climate variables performed better than those relying only on remote sensing data. Elevation-based models are effective at predicting most areas of invasion but tend to have a high over-prediction error. On the other hand, combining different data sets generally improves the models' performance, albeit not significantly (Saatchi et al., 2008). Additionally, models used in this study were successful in identifying vulnerable areas, to *L. Camara* invasion, this is because topographic variables can help to identify areas with suitable soil and moisture conditions, while Sentinel-2 bands can help to distinguish *L. Camara* from other vegetation types. The use of topographic variables and Sentinel-2 bands provide a comprehensive and accurate representation of the spatial coverage of *L. Camara*, which can be useful for invasive species management and conservation planning.

The models developed in the study had AUC values above 0.80, this indicates that both models were successful in predicting areas which are vulnerable to *L. Camara* invasion. Zhang and Foody (1998) used Sentinel to map the spatial coverage of invasive *Spartina alterniflora* in the Yangtze River estuary in China obtained an AUC of 0.87. Another study by Zhang and Foody (1998) used Landsat and the Shuttle Radar Topography Mission (SRTM) to investigate the spatial patterns and drivers of invasive Amaranthus retroflexus in the Loess Plateau of China. The authors used maximum entropy modelling (MaxEnt) to analyze the relationships between invasive plant distribution and environmental variables, such as topography and climate and obtained an AUC of 0.89. Finally, a study by Singh and Kushwaha (2013) used MaxEnt modelling to analyze the spatial patterns of *L. Camara* in the Rajaji National Park in India. The authors used environmental variables, such as topography, climate, and soil, to predict the potential suitable areas for *L. Camara* in the study area and obtained an AUC of 0.93. Overall, remote sensing and machine models have become popular tools for predicting and mapping the spatial distribution of invasive species. These models often incorporate multiple data sources, including satellite imagery, environmental data, to improve prediction accuracy and understand the drivers of invasion.

4.4. Performance of sentinel 2, random forest and MaxEnt in invasive species modelling

The results showed an overall accuracy of 84.63% and a Kappa of 0.78. This indicates the potential of RF in mapping invasive species with remote sensing data. The RF algorithm's ability to handle large datasets, multispectral data, and multicollinear datasets makes it robust in identifying and discriminating *L Camara* and other LULC types (Gil et al., 2011). The RF technique uses bagging to generate an ensemble of trees trained on multiple data subsets and randomly resampled without replacement to improve classifier stability and classification accuracy (Rodgers and Parker, 2003). Another advantage of the RF ensemble is that it results in high variance and minimal bias, and low generalization errors (Berhane et al., 2018). Several studies, including Dixon and Candade (2008), Pal and Mather (2004), and Song et al. (2012), have reported RF perform better than parametric techniques in invasive plant mapping. In addition, the data obtained from Sentinel-2 provides high spectral resolution and 13 spectral bands. Some of the bands had undergone pan-sharpening, which improves species discrimination and the accuracy of the classification (Dehkordi, 2017). While a classification accuracy of 70% or more is generally considered optimal in land use land cover mapping (Thomlinson et al., 1999), Odindi et al. (2016) noted that the 68% accuracy they achieved in classifying *L. Camara*, a plant with unclear spectral features was still useful for initial screening.

Earlier research has shown that the primary challenges in determining the spatial distribution of *L. Camara* are spectral and spatial resolutions (Tarugara et al., 2022; Khare et al., 2019; Kimothi and Dasari 2010). Sentinel-2 satellite imagery has proven to provide better imaging characteristics for mapping invasive species (Rajah et al., 2019). High-resolution sensors like Sentinel- 2 and SPOT 6, with spatial resolutions of less than 20 m, can detect the invasive *L. Camara* on a regional scale (Laso et al., 2019). Dube et al. (2014) also showed the potential of Sentinel-2 in identifying *L. Camara* invasions in Limpopo. The study reported that Sentinel-2 MSI has significant potential for distinguishing *L. Camara* from other land cover types with reasonable accuracy, with the maximum kappa statistics between 0.60 and 0.75. The spatial resolution of Sentinel-2, which is 10 m, helps to address the spectral mixing or mixed pixel problem, as noted by Immitzer et al. (2016) and Mallinis et al. (2018).

The use of bands 6, 7, (Vegetation Red Edge) and 8 (Near Infrared) in isolation produced the maximum gain but resulted in poor model performance when excluded while modelling invasive *L. Camara*. Additionally, Sentinel-2 bands 6, 7, and 8 were identified as significant variables in modelling *L. Camara* invasion. The red edge bands are critical for Sentinel-2 to produce an accurate green canopy and chlorophyll, according to Delegido et al. (2011), and this is significant for *L. Camara* prediction. The red edge is crucial for distinguishing small vegetation changes and features (Zhu et al., 2007). Vegetation red edge bands are important for the discrimination of different vegetation species (Dhau et al., 2017). Moreover, Maluleke (2019) found that Sentinel-2 band 5 (vegetation red edge) was a critical variable in discriminating *L. Camara* species in South African savannah ecosystems.

5. Conclusion

The study aimed to model the spatial patterns of *L. Camara*, as well as model areas vulnerable to invasion in Inkomati catchment in Mpumalanga, South Africa. Sentinel-2 multispectral sensor with Random Forest classifier was found to be capable of detecting and mapping *L. Camara*. Both models showed predictions that were better than random, with varying strengths depending on the variables used. However, the model that used a combination of topographic variables produced the highest AUC score. Elevation was found to be the primary variable that influenced *L. Camara* invasion. The study further shows that areas with low elevation, a gentle slope, high TPI, and low TWI are more vulnerable to *L. Camara* invasion, than other areas. The models developed in the study had excellent performance with AUC scores exceeding 0.80. However, while this study provides useful information, future research is needed to gain further insights of the invasion in different areas. This study therefore recommends the following.

- Conduct research to identify locations at risk of *L. Camara* invasion by the integration of climatic and soil data for improved invasion modelling,
- Establish water loss from *L. Camara* to aid in determining which areas are most affected by the species, from a hydrological perspective, especially for water scarce regions of sub-Saharan Africa,
- Conduct long-term monitoring, on a seasonal basis in modelling *L. Camara* over large areas.

CRediT authorship contribution statement

Vuyelwa Emmaculate Mtyobila: Data curation, Writing – review & editing, Conceptualization, Formal analysis, Validation, Visualization, Writing – original draft. **Cletah Shoko:** Conceptualization, Data curation, Funding acquisition, Investigation, Supervision, Writing – review & editing.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgements

Authors acknowledge the European Space Agency for the provision of Sentinel 2 image used in the study.

References

Adam, E., Mureriwa, N., Newete, S., 2017. Mapping Prosopis glandulosa (mesquite) in the semi-arid environment of South Africa using high-resolution WorldView-2 imagery and machine learning classifiers. J. Arid Environ. 145 (43), 51.

Adeola, A.M., 2017. Modelling vulnerability to Parthenium hysterophorus invasion in KwaZuluNatal Province, South Africa using physical, climatic, and remotely sensed derived variables. Unpublished MSc Thesis, School of Agricultural, Earth and Environmental Sciences. University of KwaZulu-Natal, Pietermaritzburg, RSA.

Ayele, S., 2007. The impact of parthenium (Parthenium hysterophorus L.) on the range ecosystem dynamics of the Jijiga rangeland, Ethiopia. Department of animal Sciences, School of Graduate studies, Haramaya University, 134.Baars, J.R. And Neser, S., 1999. Past and present initiatives on the biological control of L. Camara (Verbenaceae) in South Africa. African Entomology Memoir 1, 21–33.

Davidson, E.A., Belk, E., Boone, R.D., 1998. Soil water content and temperature as independent or confounded factors controlling soil respiration in a temperate mixed hardwood forest. Global Change Biol. 4 (2), 217–227.

Delegido, J., Verrelst, J., Alonso, L., Moreno, J., 2011. Evaluation of sentinel-2 red-edge bands for empirical estimation of green LAI and chlorophyll content. Sensors 11, 7063–7081.

Ficetola, G.F., Thuiller, W., Miaud, C., 2007. Prediction and Validation of the Potential Global Distribution of a Problematic Alien Invasive Species—The American Bullfrog.

Iqbal, I.M., Balzter, H., Shabbir, A., 2023. Mapping L. Camara and Leucaena leucocephala in Protected areas of Pakistan: a Geo-spatial approach. Rem. Sens. 15 (4), 1020.

Kimothi, M.M., Dasari, A., 2010. Methodology to map the spread of an invasive plant (L. camara.) in forest ecosystems using Indian remote sensing satellite data. Int. J. Rem. Sens. 31 (12), 3273–3289.

Kohli, R.K., Batish, D.R., Singh, H.P., Dogra, K.S., 2006. Status, invasiveness, and environmental threats of three tropical American invasive weeds (Parthenium hysterophorus L., Ageratum conyzoides L., L. Camara L.) in India. Biol. Invasions 8, 1501–1510.

Malahlela, O.E., Adjorlolo, C., Olwoch, J.M., 2022. Mapping the spatial distribution of Lippia javanica (Burm. f.) Spreng using Sentinel-2 and SRTM-derived topographic data in malaria endemic environment. Ecol. Model. 392, 147–158.

Maluleke, X.G., 2019. Modelling and Explaining the Distribution of L. Camara in South African Savanna Ecosystems (Doctoral Dissertation).

Martins, F., Alegria, C., Artur, G., 2016. Mapping invasive alien Acacia dealbata Link using ASTER multispectral imagery: a case study in central-eastern of Portugal. Forest Systems 25 (3) e078-e078.

Maxwell, A.E., Warner, T.A., Fang, F., 2018. Implementation of machine-learning classification in remote sensing: an applied review. Int. J. Rem. Sens. 39 (9), 2784–2817.

Ndlovu, H.S., Sibanda, M., Odindi, J., Buthelezi, S., Mutanga, O., 2022. Detecting and mapping the spatial distribution of Chromoleana odorata invasions in communal areas of South Africa using Sentinel-2 multispectral remotely sensed data. Phys. Chem. Earth, Parts A/B/C 126, 103081.

Ndlovu, P., Mutanga, O., Sibanda, M., Odindi, J., Rushworth, I., 2018. Modelling potential distribution of bramble (rubus cuneifolius) using topographic, bioclimatic, and remotely sensed data in the KwaZulu-Natal Drakensberg, South Africa. Appl. Geogr. 99, 54–62.

Negi, G.C., Sharma, S., Vishvakarma, S.C., Samant, S.S., Maikhuri, R.K., Prasad, R.C., Palni, L.M., 2019. Ecology and use of L. Camara in India. Bot. Rev. 85 (2), 109–130.

Odindi, J., Adam, E., Ngubane, Z., Mutanga, O., Slotow, R., 2016. Comparison between WorldView-2 and SPOT-5 images in mapping the bracken fern using the random forest algorithm. J. Appl. Remote Sens. 8, 083527.

Oumar, Z., 2016. Assessing the utility of the SPOT 6 sensor in detecting and mapping L. Camara for a Community clearing Project in KwaZulu-Natal, South Africa. S. Afr. J. Geol. 5 (2), 214–226.

Parra, J.L., Graham, C.C., D Freile, J.F., 2004. Evaluating alternative data sets for ecological niche models of birds in the Andes. Ecography 27 (3), 350–360.

Peerbhay, K., Mutanga, O., Lottering, R., Bangamwabo, V., Ismail, R., 2016. Detecting Bug weed (Solanum mauritianum) abundance in plantation forestry using multisource remote sensing. ISPRS J. Photogrammetry Remote Sens. 121, 167–176.

Phillips, S.J., Dudík, M., 2008. Modeling of species distributions with MaxEnt: new extensions and a comprehensive evaluation. Ecography 31 (2), 161–175.

Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. Ecol. Model. 190 (3), 231–259.

Priyanka, N., Joshi, P., 2013. Modelling spatial distribution of Lantana Camara-A comparative study. Canadian Journal of Basic and Applied Sciences 1, 100–117.

Qin, Z., Zhang, J.E., Ditommaso, A., Wang, R.L., Liang, K.M., 2016. Predicting the potential distribution of L. camara. under rcp scenarios using ISI-MIP models. Clim. Change 134, 193–208.

Rai, P.K., Singh, J.S., 2020. Invasive alien plant species: their impact on environment, ecosystem services and human health. Ecol. Indicat. 111, 106020.

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.

Ricciardi, A., Cohen, J., 2007. The invasiveness of an introduced species does not predict its impact. Biol. Invasions 9, 309–315.

Richardson, D.M., Rejmánek, M., 2004. Conifers as invasive aliens: a global survey and predictive framework. Divers. Distrib. 10 (5-6), 321–331.

Rodgers III, J.C., Parker, K.C., 2003. Distribution of alien plant species in relation to human disturbance on the Georgia Sea Islands. Divers. Distrib. 9, 385–398.

Saatchi, S., Buermann, W., Ter Steege, H., Mori, S., Smith, T.B., 2008. Modeling distribution of Amazonian tree species and diversity using remote sensing measurements. Remote Sensing of Environment 112 (5), 2000–2017.

Underwood, E., Ustin, S., DiPietro, D., 2003. Mapping nonnative plants using hyperspectral imagery. Remote Sensing of Environment 86 (2), 150–161.

Zhang, J., Foody, G.M., 1998. A fuzzy classification of sub-urban land cover from remotely sensed imagery. Int. J. Rem. Sens. 19, 2721–2738.

Zhu, Li, Osbert, J. Sun, Weiguo, Sang, Zhenyu, Li, Keping, M., 2007. Predicting the spatial distribution of an invasive plant species (Eupatorium adenophorum) in China. Landsc. Ecol. 22, 1143–115.