



Exploring the capability of high-resolution satellite data in delineating the potential distribution of common invasive alien plant species in the Tshivhase Tea Estate

Fhulufhedzani Nembambula¹ · Oupa E. Malahlela¹ · Lutendo Mugwedi¹

Received: 3 June 2022 / Accepted: 7 August 2023 / Published online: 18 August 2023
© The Author(s) 2023

Abstract

Invasive alien plants (IAPs) continue to exert significant impacts on agriculture in many countries, resulting in food insecurity. IAPs reduce agricultural production through competition and parasitism with planted crops. More recently, the IAPs continue to extend their plasticity to tea plantations, especially in tropical and subtropical areas. This study thus aimed at exploring the potential of SPOT 7 and Sentinel 2 satellite data in mapping the occurrence and co-occurrence of three common IAPs *Solanum mauritianum*, *Lantana camara*, and *Chromolaena odorata* in the Tshivhase Tea Estate in Limpopo Province, South Africa. The stepwise logistic regression models were generated for *Solanum mauritianum* and *Lantana camara* occurrence as well as the observed and conditional co-occurrence probability of *S. mauritianum* (P1), *L. camara* (P2) and *C. odorata* (P3). From the remote sensing indices, the Brightness Index (BI) was significant in most SPOT 7 stepwise logistic regression models at $p < 0.05$ whereas the blue, red, and near infrared (NIR) bands and standard deviation (STD_v) variables were significant at $p < 0.05$ in most of the Sentinel 2 models. The SPOT 7 model performed better than Sentinel-2 models, thus resulting in the area under the curve (AUC) of 0.96 for the conditional co-occurrence of *S. mauritianum* (P1) and *L. camara* (P2). The Sentinel 2 model yielded an AUC of 0.83. The SPOT 7 model performed superior in mapping the conditional co-occurrence of *S. mauritianum* and *L. camara* than the Sentinel 2 model. These results suggest that high spatial resolution satellite images like SPOT 7 can delineate the potential distribution of IAPs in the tea plantation and thus assisting in management strategies geared towards IAP's elimination and control.

Keywords Invasive alien plants (IAPs) · *Solanum mauritianum* · *Lantana camara* · *Chromolaena odorata* · SPOT 7 · Sentinel 2 · Logistic regression

Introduction

Invasive alien plants (IAPs) everywhere pose a substantial threat to farming, biodiversity, and the delivery of ecosystem services (Lyimo et al., 2019). The IAPs caused great economic and ecological damage around the world (Das and Duarah, 2013). They were reported to reduce agricultural productivity through various mechanisms such as competition (for light, nutrients, and water), allelopathy, and parasitism (Bajwa et al., 2019; Fried et al., 2017). Tall IAPs are

known for shading out juvenile crops and hampering their growth (Burgos and Ortuoste, 2018), thus reducing native species recruitment.

In the tropical and sub-tropical parts of South Africa, some of the noxious IAPs include the *Solanum mauritianum* (wormwood), *Lantana camara* (Lantana) and the *Chromolaena odorata* (Triffid weed). The *S. mauritianum* is a woody species native to South America and has naturalized in Africa, Australasia, India, and islands in the Atlantic, Indian and Pacific oceans (Roe, 1972). The *S. mauritianum* is a shrub or small tree that can grow up to 4 m tall (Olckers, 1999). The *L. camara* is another woody perennial shrub native to Central and South America that is aggressive and troublesome within the natural and agricultural environments worldwide with a wide climatic tolerance enabling it to be naturalised within Australia, Africa including South Africa and India (Vardien et al., 2012). On the contrary,

✉ Fhulufhedzani Nembambula
fhulunemb@yahoo.com

¹ Department of Geography and Environmental Sciences,
Faculty of Sciences, Engineering and Agriculture, University
of Venda, P/Bag X5050, Thohoyandou 0950, South Africa

C. odorata is a perennial shrub that forms dense tangled bushes of about 1.5–2.0 m in height native to subtropical and tropical America and colonised in Asia, Africa and Oceania (Koutika and Rainey, 2010). Early detection of weed infestations will provide relevant stakeholders such as farmers to control the spread of invasive alien plant species in support of efforts geared towards food security.

The advent of high spatial resolution imagery proved in many studies to enable the delineation and mapping of IAPs, as they are capable of visually detecting the spatial distribution of IAPs through direct remote sensing methods (Huang and Asner, 2009). This approach is better than the traditional field monitoring of IAPs which is difficult and expensive (Lourenco et al., 2021). Several researchers have used high spatial resolution multispectral sensors in detecting and mapping IAPs, for example, the *Pinus* spp. was successfully mapped in mountainous regions of the Western Cape Province (South Africa), using SPOT-6 imagery (Forysyth et al., 2014). Additionally, Rajah et al., 2019 have used Sentinel-2 vegetation indices (Vis) in detecting and mapping the American Bramble (*Rubus cuneifolius*) at an accuracy of 80% at the uKhahlamba Drakensberg Park (UDP). Subsequently, this integrated model correctly predicted 87% of canopy gaps that contain *C. odorata* at a probability threshold of 0.3. PlantScope (3m) and Sentinel 2 datasets were used to map the distribution of *Lantana camara* L. and *Leucaena leucocephala* L. where *L. camara* was the common species with 23% coverage and found that PlanetScope using random forest (RF), Gaussian mixture (GMM) and support vector machine (SVM) as more accurate in detecting both invasive species rather than with Sentinel 2 (Iqbal et al., 2023). Moreover, *Serephium plumosum* (*S. plumosum*) which is an encroacher shrub that has been declared as problematic in some parts of South Africa through legislation has been mapped in Telperion and Ezemvelo nature reserves in South Africa using Sentinel-2 and random forest (RF) and support vector machine (SVM) with an overall accuracy of 97.42 and 95.48% for RF and SVM respectively (Nkhwanana et al., 2022). SPOT 6 bands and vegetation indices mapped *L. camara* on an independent test dataset using a random forest algorithm at an accuracy of 75% in a community grazing land in KwaZulu- Natal (Oumar, 2016a). Kganyago et al., 2018) also used Landsat OLI 8 and SPOT 6 in mapping the distribution and patching sizes of *Parthenium hysterophorus* in the savannah landscapes of Kwazulu-Natal in combination with SVM classification at a higher overall accuracy of 86% for SPOT 6 and 83% for OLI. The medium to high spatial resolution used in all these studies was conducted in different environments other than tea plantations.

This study thus aimed to test the capability of high-medium resolution satellite imagery in mapping the occurrence of *S. mauritianum*, *L. camara*, and *C. odorata* and their co-occurrence probability in the Tshivhase Tea Estate

using SPOT 7 and Sentinel-2 images. Knowledge of the *S. mauritianum*, *L. camara* and *C. odorata* occurrence and their extent will assist Tshivhase Tea Estate managers in controlling the infestation of the IAPs within the tea plantation by targeting and prioritising areas that are invaded by IAPs. IAPs affect the growth of tea through competition for sunlight, water, nutrients, and space thereby affecting tea quality and production, therefore it is essential to detect and monitor them for easy eradication planning.

Material

Study area

The study was undertaken at Tshivhase Tea Estate in Thulamela Municipality in the Vhembe District Municipality of the Limpopo Province of South Africa lying between 30.314: 30.367 E and –22.968: –22.994 S (Figure 1) with an extent of 577 hectare (ha) (Nesamvuni et al., 2022). The Tshivhase Tea Estate is traded by Venteco, which took over the business from British firm Sapeco after it withdrew from the country in 2004 as it was operating at a loss. The Tshivhase and Mukumbani are the two sub-estates under the Tshivhase Tea Estate. The cultivation and production of tea (*Camellia sinensis*) and packaging of bulk black and rooibos tea (*Aspalathus linearis*) are the main activities in the Tshivhase Tea Estate. The study focused only on the Tshivhase sub-estate. The area is situated in the Eastern subtropical region and is generally characterised by hot humid, sub-humid and semi-arid climates (Mulugisi, 2015). It receives much of its rainfall during summer (November to February) (Kabanda, 2004; Durowoju et al., 2019).

Image data acquisition and pre-processing

The SPOT 7 and Sentinel-2 data acquired in the spring of 2021 at level 1 were used to map the presence/absence of *S. mauritianum*, *L. camara* and *C. odorata* as well as the observed and conditional co-occurrence probability amongst the three species at the Tshivhase Tea Estate. SPOT 7 has 4 spectral bands at 6m: blue (0.455–0.525 µm), green (0.530–0.590 µm), red (0.625–0.695 µm) and near-infrared (0.760–0.890 µm) with data up to 1.5 m spatial resolution panchromatic and multispectral (Hubert-Moy et al., 2020) whereas Sentinel-2 has 13 spectral bands spanning the visible/near-infrared and shortwave infrared spectral ranges at 60 m, 20 m and 10 m spatial resolutions (Immitzer et al., 2016). The 10 m spatial resolution Sentinel-2 spectral bands include: B2 (0.49 µm), B3 (0.56 µm), B4 (0.665 µm) and B8 (0.842 µm). The SPOT 7 data were pre-processed for any geometric and radiometric

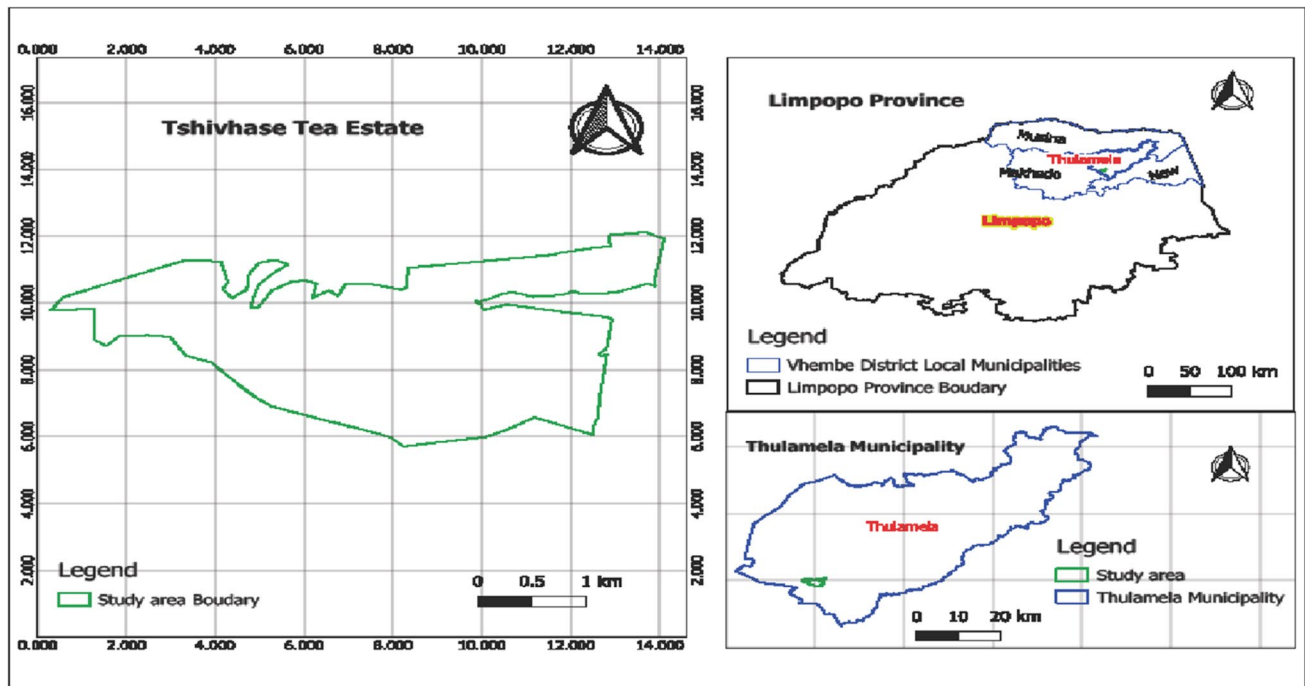


Fig. 1 Tshivhase Tea Estate, Thulamela Municipality, Limpopo Province

distortions by the South African National Space Agency (SANSA). On the other hand, Sentinel-2 data were radiometrically and geometrically corrected by the European Space Agency (ESA) (Pandzic et al., 2016). Further, Sentinel-2 was pre-processed using the open source software QGIS for atmospheric correction via the Semi-Automatic Classification Plugin (SCP) (Congedo, 2016).

Field data collection

A total of 124 samples ($n=124$) were randomly surveyed within the Tshivhase Tea Estate boundary. The locations of surveyed sites were recorded using Garmin eTrex Global Positioning System (GPS) receiver with a maximum spatial accuracy of 3 m. The presence and absence of alien invasive plant species were observed visually on the randomly sampled points for 6 days between the 1st and 8th days of October 2021. The date was chosen based on the phenological condition (green and ready for harvest) of the tea during that time. In each of the visited sites, one specimen of *S. mauritianum*, *L. camara* or *C. odorata* was considered present while no specimen observed was considered absent. In each of the visited sites, two different specimens occurring together in one site were considered co-occurrence. By observation, an infested tea plantation would have five or more specimens of the three IAPs each with a maximum height of 2 m and more.

Spectral indices

The selected spectral indices were calculated from both SPOT 7 and Sentinel-2 datasets. The following spectral indices were thus used: (i) Normalized Difference Vegetation Index (NDVI), (ii) Brightness Index (BI) and (iii) Modified Normalized Difference Water Index (MNDWI) (Table 1) were used to determine the brightness, health, and moisture status of tea plantation in the Tshivhase Tea Estate (Table 1). The NDVI was chosen as it is a successfully commonly used vegetation index in monitoring vegetation

Table 1 Spectral indices selected for the study

Vegetation Index	Formula	Reference
Normalized Difference Vegetation Index (NDVI)	$\frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + \rho_{RED}}$	(Basso et al., 2004)
Modified Normalized Difference Water Index (MNDWI)	$\frac{Green - NIR}{Green + NIR}$	Qiao et al., 2019
Brightness Index (BI)	$\frac{\sqrt{(Red^2 + Green^2)}}{2}$	Marques et al., 2020
Variance	$s^2 = \frac{\sum (x - \bar{x})^2}{n-1}$	Bhandari, 2020
Standard deviation (STDv)	$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}}$	Edward, 2020
Mean	$\bar{x} = \frac{\sum x}{n}$	Bhandari, 2020

conditions (Cropin, 2021). The MNDWI uses the green and near-infrared to extract water bodies (McFeeters, 1996) and to delineate the amount of water present in vegetation (Gao, 1996). The NDWI is useful in determining plant water stress which can have a major impact on general plant development, and crop failure resulting in lower production in agricultural areas (Factsheet, 2011). The MNDWI is not widely used in crop health monitoring and it has been used in this study to determine tea plant water stress. The Brightness Index (BI) is related to the brightness of soils influenced by soil moisture, the presence of salts, and the organic matter content of the soil surface (Escadafal, 1989). The BI was used to determine the soil moisture in the study area. In addition to the spectral indices, the standard deviation variable (STDv) and mean variable were computed from all the spectral bands for each sensor.

Method

The relationships between *S. mauritianum* presence/absence, and the remotely sensed environmental data, as well as the observed and conditional co-occurrence probability, were modelled using the stepwise logistic regression modeling (SLRM) techniques in R statistical software. The SLRM is the regression method that can be used when a response variable is binary (0/1), and the predictor variable is continuous or categorical (Sahragard and Ajorlo, 2015). Logistic regression uses the logit function to describe the relationship between response variables and predictor variables (Miller and Franklin, 2004). Logistic regression is a form of the Generalised Linear Model (GLM). The GLM is a generalisation of ordinary linear regression that allows for response variables that have error distribution models other than a normal distribution (DataCamp, 2020). The coefficients of the models were then used in QGIS to generate the distribution of the plant species map together with the co-occurrence probability map.

Model calibration

Approximately 60% of 124 total datasets ($n = 74$) were used in the stepwise logistic model calibration. The calibration dataset was used to train the model for invasive species occurrence using the R statistical software and for producing the spatial distribution maps. The SLRM was built by including all input variables in the model and the statistically insignificant variables were eliminated through the Akaike's Information Criterion (AIC) selection procedure. The logistic regression is widely used for distribution modelling of plant species in the general linear Model (Rushton et al., 2004), and it is given by the Equation (1) adapted from Higgins et al., (1999):

$$y = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n}} \quad (1)$$

where y is the probability of species occurrence, x_n is the explanatory variable, β_n is the coefficient of x_n , β_0 is the intercept and e is the exponent function of the model.

Model validation

From the standard dataset, 40% ($n = 50$) of the data was used for validating and testing the accuracy of the predictive models. The area under the curve, receiver operator characteristic (AUC ROC) measure was used for the assessment of model accuracies. The AUC statistic shows the power of the model in the distinction between presence and absence. The value of the statistics closer to 1.0 indicates superior agreement between the model and the real environment (Sahragard and Chahouki, 2015). The receiver operator characteristic (ROC) curves are used to show how a predictor compares to the true outcome (Muschelli, 2019). In ROC analyses, the predictive capabilities of a variable are commonly summarised by the area under the curve (AUC), which can be found by integrating areas under the line segments (Muschelli, 2019). AUC takes values from 0 to 1 and is considered as follows: 0 (inaccurate test), 1 (accurate test), 0.5 (no discrimination), 0.7–0.8 (acceptable), 0.8–0.9 (excellent) and more than 0.9 (outstanding) (Mandrekar, 2010).

Results

S. mauritianum occurrence

The stepwise logistic regression model derived from SPOT 7 data indicated that blue (band1), green (band 2), red (band 3), and near infrared (band4) bands were statistically significant at $p < 0.05$. Additionally, the standard deviation (STDv) was also statistically insignificant at $p > 0.05$ while the NDVI, BI and MNDWI were also significant at $p < 0.05$ with negative correlations to the presence/absence of *S. mauritianum* in Tshivhase Tea Estate (Table 2). On the other hand, the stepwise logistic regression model derived from Sentinel-2 data indicated band2 (blue), band4 (red), band 8 (NIR) and standard deviation (STDv) variables as significant at $p < 0.05$. The bands 2 and 4 and STDv yielded a positive relationship with the occurrence of *S. mauritianum* while the spectral band 8 showed a negative relationship with the presence/absence of *S. mauritianum* in the Tshivhase Tea Estate (Table 3). The SPOT 7 stepwise logistic regression yielded a higher AUC ROC of 0.93 than the Sentinel-2 model at 0.8 (Figure 2).

Table 2 The results of *S. mauritanum* occurrence stepwise logistic regression model derived from SPOT 7

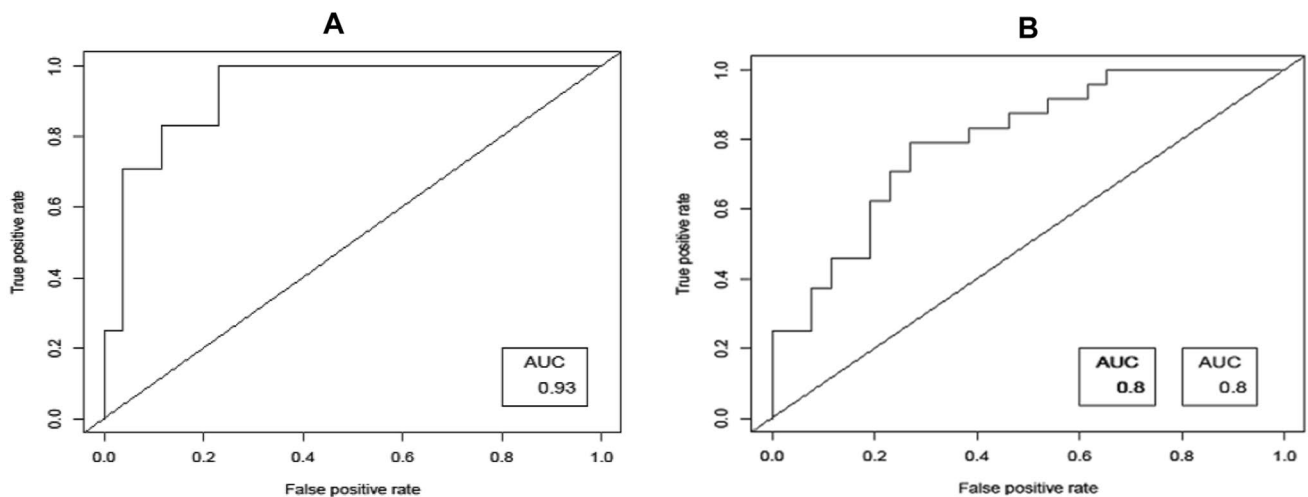
Data source	Predictor	Estimate	Standard error	Z value	P value
Final Model (SPOT)	(Intercept)	752.56	340.83	2.21	0.03*
	Band 1 (blue)	2.76	1.60	1.72	0.09.
	Band 2 (green)	1.04	0.72	1.45	0.15
	Band 3 (red)	2.92	1.54	1.90	0.06.
	Band 4 (near infrared)	−6.33	3.63	−1.74	0.08.
	Standard deviation (STDv)	12.59	7.24	1.74	0.08.
	NDVI	−1648.06	835.35	−1.97	0.05*
	BI	−1401.64	621.03	−2.26	0.02*
	MNDWI	−1939.71	961.16	−2.02	0.04*

Significance codes: °0.1; *0.05; **0.01

Table 3 The results of *S. mauritanum* occurrence stepwise logistic regression model derived from Sentinel 2

Data source	Predictor	Estimate	Standard error	Z value	P value
	(Intercept)	−4.39	8.63	−0.51	0.61
	band 2(blue)	719.98	215.74	3.34	0.00***
	band 4(red)	1156.90	308.41	3.75	0.00***
	band 8(near infrared)	−1933.80	485.30	−3.99	0.00***
	Standard deviation (STDv)	4110.43	1023.71	4.02	0.00***

Significance codes: *0.05; **0.01; ***0.001

**Fig. 2** SPOT 7 (A) and Sentinel-2 (B) *S. mauritanum* occurrence Logistic Regression model ROC AUC curve

L. camara occurrence

The stepwise logistic regression model derived from SPOT 7 variables indicated red (band 3), standard deviation (STDv), variance, BI and MNDWI as significant at $p < 0.05$ with red (band 3) and variance having a positive relationship with the occurrence of *L. camara* and standard deviation (STDv). Additionally, the BI and MNDWI yielded a negative correlation with the presence and absence of *L.*

camara (Table 4). On the other hand, the regression model derived from Sentinel-2 variables indicated a blue band as statistically significant at $p < 0.05$ with a negative correlation with the presence/absence of *L. camara*. The red, near infrared, STDv and BI were not significant (Table 5). The stepwise logistic regression model derived from the SPOT 7 data model yielded a performance prediction of 0.87 AUC ROC whereas Sentinel-2 produced a prediction performance of 0.63 (Figure 3).

Table 4 The results of *L. camara* occurrence stepwise logistic regression model derived from SPOT 7

Data source	Predictor	Estimate	Standard error	Z value	P value
SPOT 7	(Intercept)	−100.70	42.32	−2.38	0.01729*
	band 3(red)	0.55	0.17	3.206	0.00135**
	Standard deviation (STDv)	−0.65	0.24	−2.753	0.00591**
	Variance	0.00	0.00	2.074	0.03808*
	BI	−412.00	130.60	−3.154	0.00161**
	MNDWI	−797.90	253.40	−3.148	0.00164**

Significance codes: °0.1; *0.05; **0.01; ***0.001

Observed co-occurrence probability

The model derived from SPOT 7 for the observed co-occurrence probability of *S. mauritianum* (P1) and *L.*

camara (P2) indicated band 1 as a statistical significance at $p < 0.05$ with a positive relationship to the occurrence of P1 and P2 IAP species whereas red (band 3), near infrared (band 4), variance and BI as insignificant at $p > 0.05$ (Table 6). Other variables such as soil condition and Digital Elevation Model (DEM) could correlate with the presence/absence of *L. camara* in the Tshivhase Tea Estate. The stepwise logistic regression derived from Sentinel 2 data indicated blue (band 2), red (band 4), near infrared (B8), standard deviation (STDv), and BI as significant at $p < 0.05$ with band 2, band 4, and STDv shows a negative correlation towards the co-occurrence of P1 (*S. mauritianum*) and P2 (*L. camara*) (Table 7). The variance was insignificant at $p > 0.05$ (Table 7). The SPOT 7 model produced a higher AUC ROC performance of 0.91 in predicting the co-occurrence of *S. mauritianum* and *L. camara* in the Tshivhase Tea Estate whereas sentinel 2 produced an AUC ROC of 0.82 (Figure 4).

The observed co-occurrence probability model of *S. mauritianum* (P1) and *C. odorata* (P3) for Sentinel 2

Table 5 The results of *L. camara* occurrence stepwise logistic regression model derived from Sentinel 2

Data source	Predictor	Estimate	Standard error	Z value	P value
Sentinel 2	(Intercept)	−111.26	81.16	−1.37	0.17
	band 2(blue)	−1523.95	648.30	−2.35	0.02*
	band 4(red)	−5688.89	3121.29	−1.82	0.07.
	Band 8(near infrared)	5830.76	3121.20	1.87	0.06.
	Standard deviation(STDv)	−12,025.16	6448.30	−1.87	0.06.
	BI	370.91	191.63	1.94	0.05.

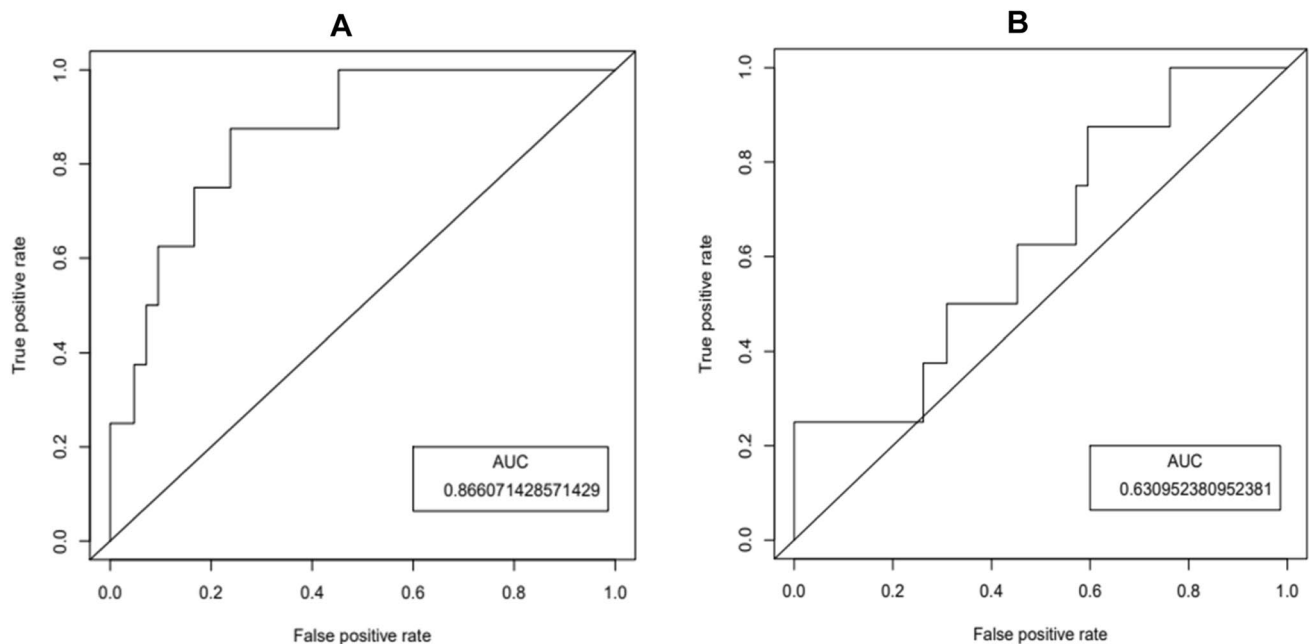
**Fig. 3** SPOT 7 (A) and Sentinel-2 (B) *L. camara* occurrence Logistic Regression model ROC AUC curve

Table 6 The results of observed co-occurrence probability stepwise logistic regression model derived from SPOT 7

Data source	Predictor	Estimate	Standard error	Z value	P value
SPOT 7	(Intercept)	−70.94	47.02	−1.51	0.13
	band 1(blue)	0.15	0.07	2.25	0.02*
	band 3(red)	0.93	0.58	1.61	0.11
	band 4(near infrared)	−0.41	0.26	−1.58	0.11
	variance	0.00	0.00	1.38	0.17
	BI	−575.00	342.60	−1.68	0.09
	MNDWI	−979.20	560.60	−1.75	0.08

Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Table 7 The results of observed co-occurrence probability stepwise logistic regression model derived from Sentinel 2

Data source	Predictor	Estimate	Standard error	Z value	P value
Sentinel 2	(Intercept)	−233.50	127.20	−1.84	0.07.
	band 2(blue)	−2759.50	1265.10	−2.18	0.03*
	band 4(red)	−11899.20	5947.80	−2.00	0.05*
	band 8(near infrared)	12146.40	6029.70	2.01	0.04*
	Standard deviation(STDv)	−25787.00	12852.10	−2.01	0.04*
	variance	4302.40	2734.30	1.57	0.12
	BI	752.70	365.50	2.06	0.04*

Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

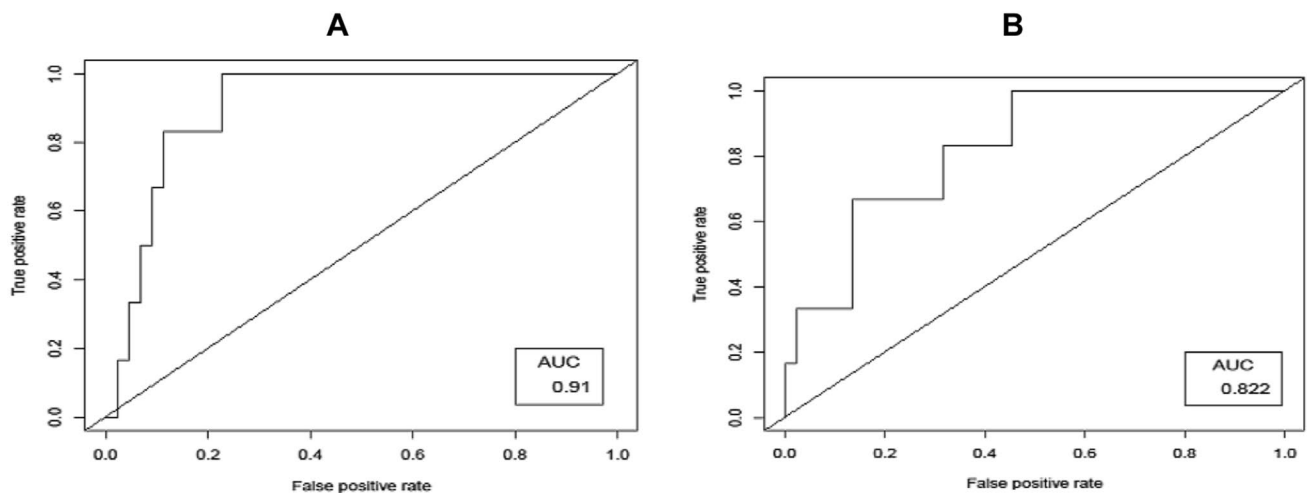


Fig. 4 SPOT 7 (A) and Sentinel-2 (B) observed co-occurrence of *S. mauritanum* (P1) and *C. odorata* (P3) Logistic Regression model ROC AUC curve

could not be validated as it was invalid for binary logistic models. On the actual presence/absence data in validation data, there was no co-occurrence of P1 and P3 species.

Conditional co-occurrence probability

The stepwise logistic regression derived from SPOT 7 data indicated only blue (band 1) as significant and red (band 3), near infrared (Band 4), variance, BI, and MNDWI as

non-significant in predicting the co-occurrence of *S. mauritanum* and *L. camara* given that *C. odorata* is not occurring (Table 8). The stepwise logistic regression derived from sentinel 2 data on the other hand indicated all the spectral bands, STDv, mean, variance and BI as non-significant at $p > 0.05$ (Table 9). This means that these variables do not influence the co-occurrence of *S. mauritanum* and *L. camara* in the Tshivhase Tea Estate. Other variables such as soil condition, climate, and DEM could have an

influence over the co-occurrence of *S. mauritianum* and *L. camara* in the condition where *C. odorata* does not occur. The SPOT 7 stepwise logistic regression of conditional co-occurrence probability model derived from SPOT 7 data yielded an AUC ROC of 0.96 whereas the Sentinel 2 model yielded 0.83 (Figure 5).

The conditional co-occurrence probability model of both SPOT 7 and Sentinel-2 was insignificant in predicting the co-occurrence of *S. mauritianum* and *C. odorata* given that *L. camara* does not occur. There was no probability that *S. mauritianum* and *C. odorata* co-occur in one area in the Tshivhase Tea Estate.

Table 8 The results of conditional co-occurrence probability (PIP2) stepwise logistic regression model derived from SPOT 7

Data source	Predictor	Estimate	Standard error	Z value	P value
SPOT 7	(Intercept)	−68.67	46.28	−1.48	0.14
	band 1(blue)	0.15	0.07	2.09	0.04*
	band 3(red)	1.00	0.63	1.60	0.11
	band 4(near infrared)	−0.43	0.28	−1.54	0.12
	variance	0.00	0.00	1.39	0.16
	BI	−597.30	369.70	−1.62	0.11
	MNDWI	−994.20	597.70	−1.66	0.10.

Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘.’ 1

Table 9 The results of conditional co-occurrence probability (PIP2) stepwise logistic regression model derived from Sentinel 2

Data source	Predictor	Estimate	SE	Z value	P value
Sentinel 2	(Intercept)	−237.70	134.40	−1.77	0.08.
	B2(Blue)	−2997.50	1369.20	−2.19	0.03*
	B4(Red)	−12,385.50	6329.20	−1.96	0.05.
	B8(Near Infrared)	12,748.00	6437.70	1.98	0.05*
	Standard deviation(STDv)	−27,027.60	13,725.60	−1.97	0.05*
	Variance	4265.30	2950.30	1.45	0.15
	BI	780.70	388.50	2.01	0.04*

Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘.’ 1

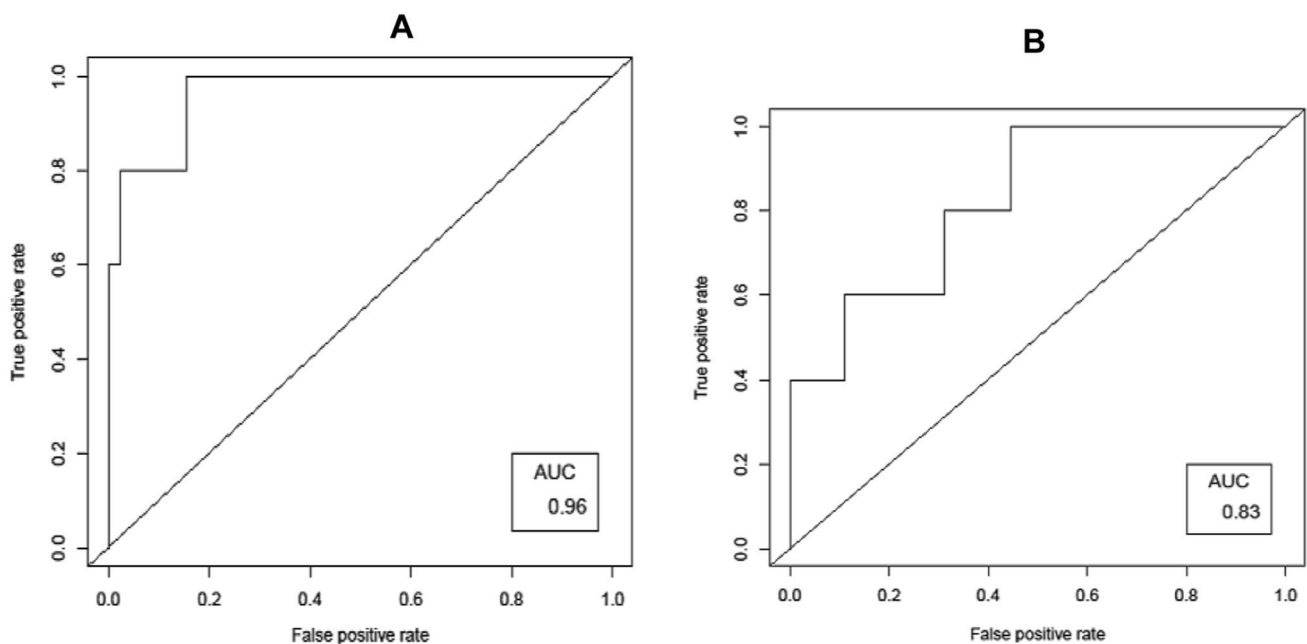


Fig. 5 SPOT 7 (A) and Sentinel-2 (B) conditional co-occurrence Logistic Regression model of P1 and P2 ROC AUC curve

Discussion

S. mauritianum occurrence

The stepwise logistic regression model derived from SPOT 7 indicated NDVI, BI, and MNDWI as significant with a negative relationship to the occurrence of *S. mauritianum*. This means that tea health, soil moisture, and moisture condition do not influence the occurrence of *S. mauritianum* in the Tshivhase Tea Estate. On the other hand, stepwise logistic regression derived from Sentinel-2 showed blue (band 2), red (band 4) and standard deviation (STDv) influencing the occurrence of *S. mauritianum*.

The current SPOT 7 model in this study has exhibited a higher AUC (AUC = 0.93) than Sentinel 2 with 0.8 AUC. AUC is very important for understanding the accuracy and reliability of the predictive model. SPOT 7 models performed better than the Sentinel 2 model because of its high spatial resolution. Many studies successfully used SPOT data in detecting and mapping IAPs (Odindi et al., 2014; Schmidt et al., 2010). For instance, Odindi et al., 2014 mapped bracken fern using WorldView-2 and SPOT-5 images with an overall accuracy of 84.72% and 72.22% respectively use the random forest algorithm. In this study, the stepwise logistic regression model derived from a high spatial resolution image was superior in mapping the occurrence of *S. mauritianum*. This is in support of the previous

studies that have successfully used SPOT data in monitoring and mapping IAPs. Most of these studies were conducted in a different environment other than the tea plantation such as the forests, community grazing land, and mountainous areas. The high spatial configurations and temporal resolution of SPOT data may have improved the identification and monitoring of IAPs in these different environments.

Figure 6 shows the predicted spatial distribution of *S. mauritianum* in the Tshivhase Tea Estate where the blue colour represents where the *S. mauritianum* occurs and the red colour represents where the species do not occur. SPOT 7 has successfully mapped the distribution of *S. mauritianum* over the Sentinel 2 map and conforms to what is on the ground (Figure 6). The occurrence of *S. mauritianum* could have been improved by increasing the sampled points within the tea estate. The distribution pattern of *S. mauritianum* will assist the Tshivhase Tea Estate managers in better planning for eradication. From the field observation, monkeys were feeding on the *S. mauritianum* fruits, distributing seeds all over the tea plantation Fig. 7.

L. camara occurrence

The stepwise logistic regression model derived from SPOT 7 indicated B3 and variance as significant with a positive relationship towards the occurrence of *L. camara* in the study area whereas Green (B2) was significant with a

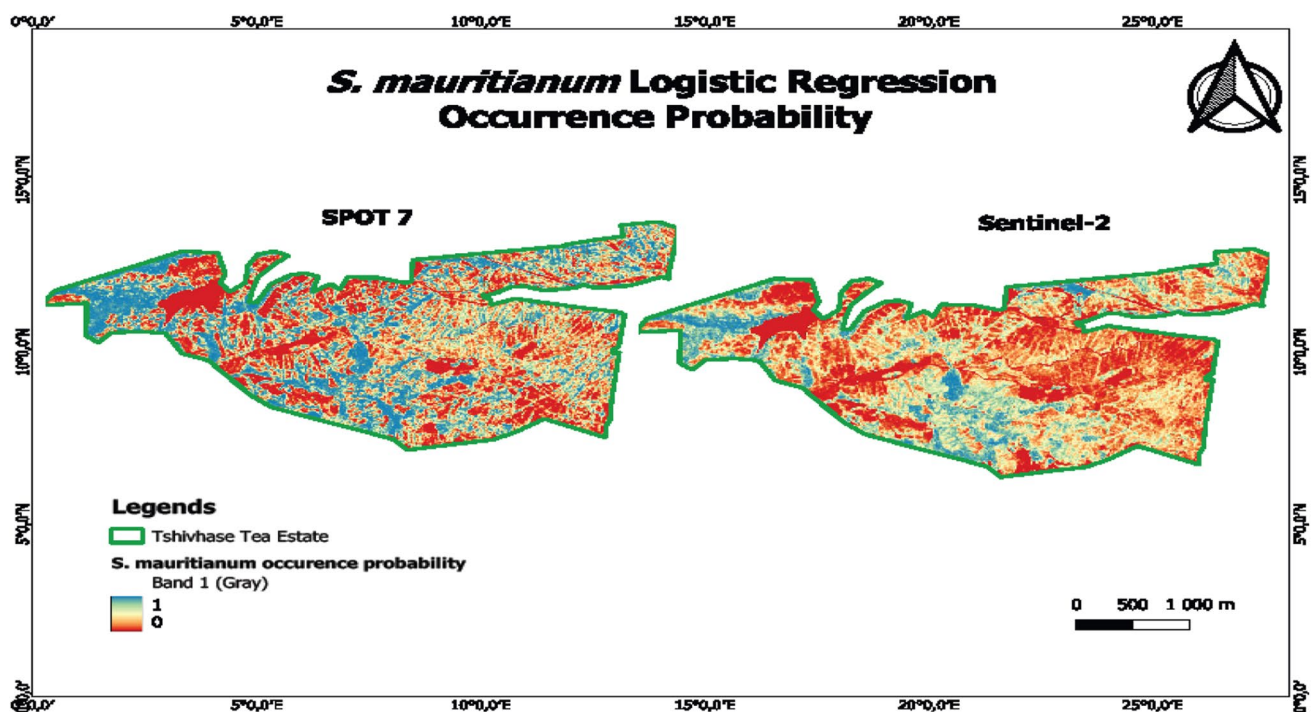


Fig. 6 *S. mauritianum* occurrence stepwise logistic regression model map for SPOT 7 and Sentinel 2

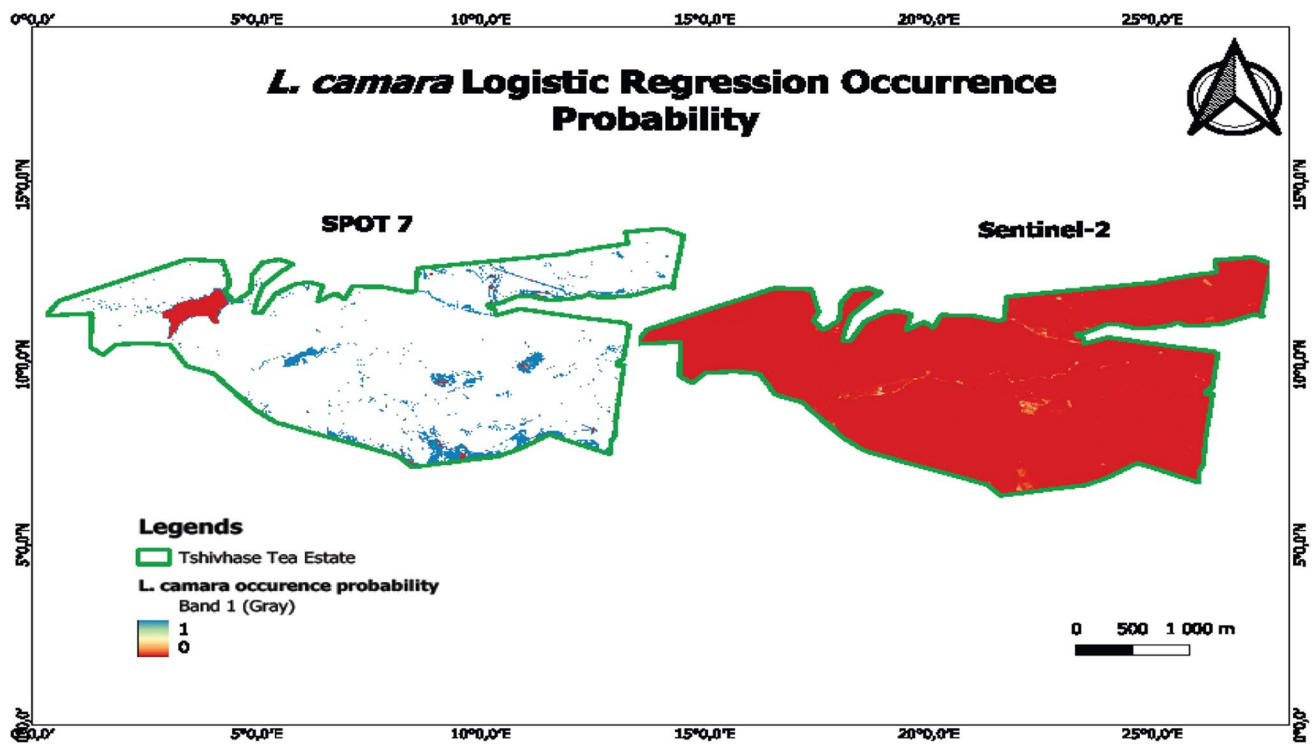


Fig. 7 *L. camara* occurrence stepwise logistic regression model map for SPOT 7 and Sentinel 2

negative relationship to the occurrence of *L. camaradata* on the Sentinel 2 stepwise logistic regression model. The stepwise logistic regression model derived from SPOT 7 yielded a performance prediction of 0.87 whereas Sentinel 2 produced a prediction performance of 0.61 (Figure 4). This means that SPOT 7 was able to predict the occurrence of *L. camara* in the Tshivhase Tea Estate than Sentinel 2 because of its high spatial resolution property. The potential of satellite remote sensing in weed detection and mapping in South Africa using readily available multispectral data has been proven in the assessment of the SPOT 6 sensor in detecting and mapping *L. camara* for a community clearing project in the KwaZulu-Natal study (Oumar, 2016). This study is congruent with Omar's (2016) study as it has also mapped the occurrence of *L. camara* using SPOT 7 successfully. Sentinel 2 was also used successfully in mapping the distribution of IAPs in a water-limited catchment with a higher overall accuracy of 71% (Mtengwana, 2020). The map in Figure 7 shows the occurrence of *L. camara* in the Tshivhase Tea Estate derived from SPOT 7 representing where the species occurs in blue colour. The Sentinel 2 map indicated that there is no occurrence of *L. camara* with a red colour and therefore SPOT 7 still stands in mapping the occurrence of *L. camara*.

Observed co-occurrence probability

The stepwise logistic regression model derived from SPOT 7 data indicated band 1 as significant and with a positive relationship towards the co-occurrence of *S. mauritianum* (P1) and *L. camara* (P2) in one area. The Sentinel 2 stepwise logistic regression model indicated band 8 and BI as significant and with a positive relationship towards the co-occurrence of P1 and P2 species whereas blue (band 2), red (band 4), standard deviation (STDv) as significant with a negative relationship towards the co-occurrence of these two species. The SPOT 7 stepwise logistic regression model produced a higher AUC ROC of 0.91 in predicting the co-occurrence of P1 and P2 species than the Sentinel 2 model with 0.82. For future studies, high spatial resolution images should be used with machine learning algorithms to delineate various IAPs species occurring in one area.

The maps below show the observed co-occurrence probability of *S. mauritianum* and *L. camara* (Figure 8). The SPOT 7 map indicates *S. mauritianum* and *L. camara* occurring together in one area along the open areas whereas the Sentinel 2 map shows no co-occurrence of these two species.

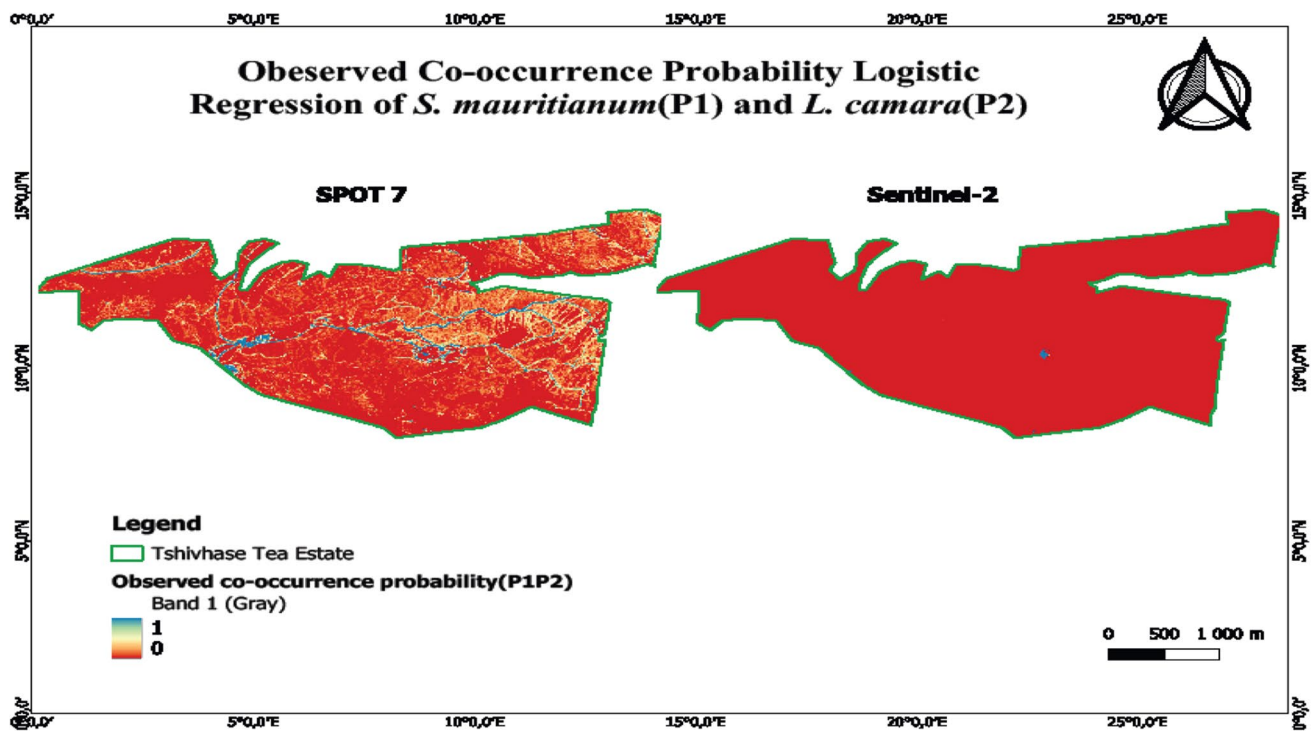


Fig. 8 Observed co-occurrence stepwise logistic regression model map for SPOT 7 and Sentinel 2

Conditional co-occurrence probability

The stepwise logistic regression derived from SPOT 7 data indicated blue (band 1) as significant with a positive relationship towards the co-occurrence of *S. mauritianum* (P1) and *L. camara* (P2) co-occurring in one area together give that *C. odorata* (P3) does not occur whereas the other variable was insignificant. The conditional co-occurrence probability model of *S. mauritianum* (P1) and *L. camara* (P2) for SPOT 7 yielded ROC AUC of 0.96 whereas the Sentinel 2 model yielded 0.83. This means that SPOT 7 was good in predicting the occurrence of *S. mauritianum* and *L. camara* in one area together given that *C. odorata* is not occurring. For future studies, conditional co-occurrence probability should be applied to a larger area using hyperspectral satellite data with the capability of delineating different IAPs under the canopy.

Conclusion

Tea plantations are infested with IAPs species affecting their growth and health and therefore innovative, effective, dependable, and easily accessible measures are needed to monitor the spread of IAPs species within the

tea plantation. This study aimed at determining the potential of the medium to high spatial resolution satellite data in mapping the existing occurrence and the co-occurrence of IAPs species within the Tshivhase Tea Estate. The combination of logistic regression and remote sensing variables was significant in mapping the prediction of *S. mauritianum*, *L. camara* and *C. odorata* especially when using a higher spatial resolution satellite image like SPOT 7. The observational matrix was also successful in mapping the co-occurrence between the IAPs species and therefore they can assist the tea plantation managers with information about the IAPs species that co-occur within the tea plantation and plan for clearance thereof. Using logistic regression and remote sensing variables, the location of these IAPs species will be easily identifiable providing the opportunity to directly target areas infested by these species saving time for walking all over the tea plantation locating areas invaded by invasive alien plant species. This study introduced the use of an observational matrix in combination with satellite data in mapping IAPs species that occur together in one area and IAP species that do not occur when the others occur together. This will give tea farmers an opportunity to understand IAPs species community within the tea plantation and plan for their clearance thereof.

Acknowledgements Appreciation goes to the University of Venda and SANSA. The authors would further like to thank Ms. Kgaogelo Mogano from Agricultural Research Council: Natural resource and Engineering for dedicating her time and effort in assisting with field data collection.

Author contributions Research Investigation, Methodology, Data Curation, Analysis, Interpretation of Results, and Writing—Original Draft, F.N.; Resources and Editing, O.M.; Research Conceptualization, Editing and Supervision O.A and L.M. All authors have read and agreed to the published version of the manuscript.

Funding Open access funding provided by University of Venda. The South African National Space Agency (SANSA) funded this study.

Data Availability The raw data supporting the findings of this study are available upon reasonable request from the corresponding author. Satellite images were sourced from USGS earth explorer site and SANSA.

Declarations

Conflict of interest The authors declare no competing interests.

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

References

- Bajwa AA, Farooq M, Nawaz A, Yadav L, Chauhan BS, Adkins S (2019) Impact of invasive plant species on the livelihoods of farming households: evidence from *Parthenium hysterophorus* invasion in rural Punjab, Pakistan. *Biol Invasions* 21(11):3285–3304
- Basso B, Cammarano D, De Vita P (2004) Remotely sensed vegetation indices: theory and applications for crop management. *Rivista Italiana di Agrometeorologia* 1(2004):36–53
- Bhandari P (2020) How to find the mean | definition, examples & calculator. Scribb. <https://www.scribb.com/statistics/mean/>
- Burgos NR, Ortuoste JD (2018) *Weed management in natural rubber. weed control: sustainability, hazards, and risks in cropping systems worldwide*, Chapter 26, 1st edn. CRC Press, Boca Raton, pp 486–504
- Congedo L (2016) Semi-automatic classification plugin documentation. <https://doi.org/10.13140/RG.2.2.29474.02242/1>
- Cropin (2021) NDVI and its practical uses in agriculture. <https://www.cropin.com/blogs/ndvi-in-agriculture>
- Das K, Duarah P (2013) Invasive alien plant species in the roadside areas of Jorhat, Assam: their harmful effects and beneficial uses. *Int J Eng Res Appl* 3(5):353–358
- DataCamp (2020) GLM in R: generalized linear model. <https://www.datacamp.com/tutorial/generalized-linear-models>
- Durowoju OS, Odiyo JO, Ekosse GIE (2019) Determination of isotopic composition of rainwater to generate local meteoric water line in Thohoyandou, Limpopo Province, South Africa. *Water SA* 45(2):183–189
- Edward F (2020) Standard deviation calculator. <https://www.calculator-soup.com/calculators/statistics/standard-deviation-calculator.php>
- Escadafal R (1989) Remote sensing of arid soil surface color with Landsat thematic mapper. *Adv Space Res* 9(1):159–163
- Factsheet (2011) NDWI: Normalized Difference Water Index. Version 1. https://edo.jrc.ec.europa.eu/documents/factsheets/factsheet_ndwi.pdf
- Forsyth AT, Gibson LA, Turner AA (2014) Assessment of SPOT 6 imagery for mapping the invasive alien plant species *Pinus* spp. In: A Mountainous Area of the Western Cape. http://www.africageoprocceedings.org.za/wp-content/uploads/2014/06/166_Forsyth_Gibson_Turner.pdf
- Fried G, Chauvel B, Reynaud P, Sache I (2017) Decreases in crop production by non-native weeds, pests, and pathogens. *Invading Nature - Springer Series in Invasion Ecology* 12:83–101
- Gao B (1996) NDWI - a normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens of Environ* 58(3):257–266
- Higgins SI, Richardson DM, Cowling RM, Trinder-Smith TH (1999) Predicting the landscape-scale distribution of alien plants and their threat to plant diversity. *Conserv Biol* 13:303–313 <https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi/resolutions/spatial>
- Huang CY, Asner GP (2009) Applications of Remote Sensing to Alien Invasive Plant Studies. *Sensors* 9(6):4869–4889. <https://doi.org/10.3390/s90604869>
- Hubert-Moy L, Fabre E, Rapinel S (2020) Contribution of SPOT-7 multi-temporal imagery for mapping wetland vegetation. *Eur J Remote Sens* 53(1):201–210. <https://doi.org/10.1080/22797254.2020.1795727>
- Immitzer M, Vuolo F, Atzberger C (2016) First experience with Sentinel-2 data for crop and tree species classifications in Central Europe. *Remote Sens* 8:166
- Kabanda TA (2004) Climatology of long-term drought in northern region of Limpopo province'. PhD thesis, Dept. of Geography and Geo-information Sciences, University of Venda
- Kganyago M, Odindi J, Adjorlolo C, Mhangara P (2018) Evaluating the capability of Landsat 8 OLI and SPOT 6 for discriminating invasive alien species in the African Savanna landscape. *Int J Appl Earth Obs and Geoinf* 67:10–19. <https://doi.org/10.1016/j.jag.2017.12.008>
- Koutika LS, Rainey HS (2010) *Chromolaena odorata* in different ecosystems: Weed or fallow plant? *Appl Ecol Environ Res* 8(2):131–142
- Lourenço P, Teodorode AC, Gonçalves J, Honrado JP, Cunha M, Silero N (2021) Assessing the performance of different OBIA software approaches for mapping invasive alien plants along roads with remote sensing data. *Int J Appl Earth Obs Geoinf* 95:102263 <https://doi.org/10.1097/JTO.0b013e3181ec173d>
- Lyimo P, Macrice S, Emily C, Munishi P, Paul C, Busingo A, Masaka K, Kitasho N, Justine S (2019) Invasive Alien Plant Species in Tanzania. Sokoine University of Agriculture, Department of Ecosystems and Conservation
- Mandrekhar JN (2010) Receiver Operating Characteristic Curve in Diagnostic Test Assessment. *J Thorac Oncol* 5(9):1315–1316. <https://doi.org/10.1097/JTO.0b013e3181ec173d>
- Marqués MJ, Alvarez A, Carral P, Sastre B, Bienes R (2020) The use of remote sensing to detect the consequences of erosion in gypsiferous soils. *Int Soil Water Conserv Res* 8:383–392
- McFeeters S (1996) The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *Int J Remote Sens* 17(7):1425–1432
- Miller D, Mohanty BP, Jacobs JM, Hsu EC (2004) SMEX02: field scale variability, time stability and similarity of soil moisture. *Remote Sens Environ* 92:436–446

- Mtengwana B, Dube T, Mkunanya YP, Mazvimavi D (2020) Use of multispectral satellite datasets to improve ecological understanding of the distribution of Invasive Alien Plants in a water-limited catchment, South Africa. *Afr J Ecol* 58(4):709–718
- Mulugisi A (2015) The impacts of heavy rains on the vegetation cover in the Limpopo Province of South Africa. Dissertation submitted in fulfillment of the requirements for the degree of Master of Science in Environmental Science at the Mafikeng Campus North-West University. https://repository.nwu.ac.za/bitstream/handle/10394/17385/Mulugisi_A.pdf?sequence=1&isAllowed=y
- Muschelli J (2019) ROC and AUC with a Binary Predictor: a Potentially Misleading Metric. https://www.researchgate.net/publication/331700624_ROC_and_AUC_with_a_Binary_Predictor_a_Potentially_Misleading_Metric
- Nesamvuni AE, Bokosi J, Tshikolomo KA, Mpandeli NS, Nesamvuni C (2022) Tea value chains viability in Limpopo Province of South Africa: a cost-benefit analysis. In: Sustainable Agricultural Value Chain. Intechopen. <https://doi.org/10.5772/intechopen.102081>
- Nkhwanana N, Adam E, Ramoelo A (2022) Assessing the utility of Sentinel-2 MSI in mapping an encroaching *Serephium plumosum* in South African rangeland. *Appl Geomat* 14:435–449
- Odindi J, Adam E, Ngubane Z, Mutanga O, Slotow R (2014) Comparison between WorldView2 and SPOT-5 images in mapping the bracken fern using the random forest algorithm. *J Appl Remote Sens* 8:083527
- Olckers T (1999) Biological control of *Solanum mauritianum* Scopoli (Solanaceae) in South Africa: a review of candidate agents, progress, and future prospects. *Afr Entomol Memoir* 1:65–73
- Oumar Z (2016) Assessing the utility of the SPOT 6 sensor in detecting and mapping *Lantana camara* for a community clearing project in KwaZulu-Natal, South Africa. *S Afr J Geomat* 5(2):214–226
- Pandzic M, Mihajlovic D, Pandzic J, Pfeifer N (2016) Assessment of the geometric quality of Sentinel-2 data. In: *Int Arch Photogramm Remote Sens Spat Inf Sci* 41:489–494
- Qbal IM, Balzter H, Bareen F, Shabbir A (2023) Mapping *Lantana camara* and *Leucaena leucocephala* in protected areas of Pakistan: a geo-spatial approach. *Remote Sens* 15(4):1020. <https://doi.org/10.3390/rs15041020>
- Qiao B, Zhu L, Yang R (2019) Temporal-spatial differences in lake water storage changes and their links to climate change throughout the Tibetan Plateau. *Remote Sens Environ* 222:232–243
- Rajah P, Odindi J, Mutanga O (2019) 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. *Nature Conser* 35:41–61
- Roe KE (1972) A revision of *Solanum* sect. *Brevantherum* (Solanaceae). *Brittonia* 24:239–278
- Rushton SP, Ormerod SJ, Kerby G (2004) New paradigms for modelling species distributions? *J Appl Ecol* 41(2):193–200
- Sahragard PH, Chahouki ZMA (2015) An evaluation of predictive habitat models performance of plant species in Hoze-soltan rangelands of Qom province. *Ecol Model* 309–310:64–71
- Schmidt M, Witte C (2010) Monitoring aquatic weeds in a river system using SPOT 5 satellite imagery. *J Appl Remote Sens* 4:043528
- Vardien W, Richardson DM, Foxcroft LC, Thompson GD, Wilson JRU, Le Roux JJ (2012) Invasion dynamics of *Lantana camara* L. (*sensu lato*) in South Africa. *S. Afr. J. Bot* 81:81–94

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