Mapping the probability of occurrence of invasive *Chromolaena odorata* in subtropical forest gaps using environmental and remote sensing data

Oupa E. Malahlela* [1, 2], Moses A. Cho [1, 2], Onisimo Mutanga [2]

[1] *Earth Observation Research Group, Natural Resources and Environment, Council for Scientific and Industrial Research, Pretoria 0001, South Africa*
[2] *Geography Department, University of KwaZulu-Natal, Pietermaritzburg Campus, Scottsville 3209, South Africa*

*Corresponding author. Current address: P O Box 395, Pretoria, 0001, South Africa. Tel: +27 128412233. Email address: OMalahlela@csir.co.za
M. A. Cho. Current address: P O Box 395, Pretoria, 0001, South Africa. Tel: +27 128413669. Email address: MCho@csir.co.za
O. Mutanga. Current address: P. Bag X01, Scottsville 3209, Pietermaritzburg, South Africa. Email address: MutangaO@ukzn.ac.za

Abstract

Globally, subtropical forests are rich in biodiversity. However, the native biodiversity of these forests is threatened by the presence of invasive species such as *Chromolaena odorata* which thrives in forest canopy gaps. Our study explored the utility of WorldView-2 data, an 8-band high resolution (2 m) imagery for mapping the probability of *C. odorata* occurrence (presence/absence) in canopy gaps of a subtropical forest patch, the Dukuduku Forest, South Africa. An integrated modelling approach involving the WorldView-2 bands and ancillary environmental data was also assessed. The results showed a higher performance of the environmental data only model (deviance or D^2 = 0.52, $p < 0.05$, n = 77) when compared to modelling with WorldView-2 vegetation indices such as the enhanced vegetation index (EVI), simple ratio indices (SRI) and red edge normalized difference vegetation index (NDVIr) (D^2 = 0.30, $p < 0.05$, n = 77). The integrated model explained the highest presence/absence variance of *C. odorata* (D^2 =0.57, i.e. 57%). This model was used to derive probability map indicating the occurrence of invasive species in forest gaps. A 2 x 2 error matrix table and the receiver operating characteristic (ROC) curves derived from an independent validation dataset (n = 38) were used to assess the mapping accuracy. Approximately 87% of canopy gaps containing *C. odorata* were correctly predicted at probability threshold of between 0.2 and 0.3. The derived probability map of *C. odorata* occurrence will assist management in prioritizing target areas for eradication of the species.

Keywords: forest management, remote sensing, invasive species, ROC curve, mapping accuracy

Introduction

Tropical forests cover approximately 6 % of the total land surface, and are important to the survival of many faunal and floral life-forms (WWF 2013). These forests also play a major role in the global climatic system, acting as carbon sinks (Sohngen and Alig 2000) and in water and nutrient cycling (Skole and Tucker 1993; Bousquet et al. 2000). Additionally, they provide ecological services to the nearby communities in the form of harvested timber, medicinal, food and wood crafts (Balee 1989;

Cunningham 2001). However, their sustainability is threatened by the anthropogenic activities such as settlement expansion, agriculture, industrial activities, climate change and invasive species (Fourcade 1889; Geldenhuys 1989).

One of the world's most noxious alien plants is *Chromolaena odorata* (triffid weed, Fig.1.), a shrub species that is indigenous to North and Central America, and is invasive in more than 23 countries globally. The invasive *Chromolaena odorata* (here onwards referred to as *C. odorata*) has an allelopathic effect that inhibits the indigenous plant species recruitment (Goodall and Zachariades, 2002; Sahid and Sugau, 1993), by changing the chemical composition of soil underneath its canopy in indigenous forests (de Rouw 1991). *C. odorata* thrives in habitats that receive sufficient sunlight, ruderal environments, close to water courses and at forest or road margins (Joshi et al. 2006). The habitat preference is facilitated by the fact that *Chromolaena* species requires sufficient light and moisture, and as such, wide open areas serve as optimal niches for its establishment and seed dispersal (Witkowski and Wilson 2001). Its tap-root system allows for deeper penetration into the substrate which gives it a competitive advantage over the native seedlings (Awanyo et al. 2011). In tropical forests, this species invades the indigenous vegetation through forest gaps, which may be created by tree fall (Kupfer and Runkle 1996), any catastrophic event (Brokaw 1982; Whitmore 1989), or selective timber harvesting (Suarez et al. 1998). The control and eradication of these invaders requires accurate mapping of forest gaps and modelling the extent of invasion in such gaps (Le Maitre et al. 1996; Underwood et al. 2003). Mapping the extent of invasive species may assist forest managers, conservation practitioners and all the relevant stakeholders to comprehend the extent of invasion and to allocate limited resources for the species eradication programmes (Reyers 2004).

Conventionally, identification of invasive species is done using ground-based surveys (Buckland et al. 1996; Scott et al. 2002). This involves identifying species at habitats that are accessible, usually at the forest edges and in the proximity to roads and paths (Edwards et al. 2007). Although the collected environmental data (such as distance from roads and forest edges) are still useful in species prediction (Hirzel and Guisan 2002), the success of using environmental data collected through traditional field surveys is hampered by the amount of time and effort required for collection (Gu and Swihart 2004; Margules and Pressey 2000). Additionally, a major drawback of employing field surveys is that they are inefficient in larger areas and in areas with less accessible terrain (Turner et al. 2003), while it is impossible to visit all forest gaps in indigenous tropical forests. To mitigate this problem, remote sensing technology is increasingly being employed as a rapid and cost-effective alternative for mapping invasive species in their habitats (Underwood et al. 2003; Joshi et al. 2006). Remote sensing technology provides spatial vegetation cover over large geographical areas. This technology has been successfully used for actual canopy cover mapping of invasive species that dominate the canopy (Asner et al. 2008; Harding and Bate 1991). In instances where the invasive species' spectral signature does not dominate the canopy, the probabilistic mapping approach has been adopted (Joshi et al. 2006; Laba et al. 2004).

Few studies have been conducted to map the probability of occurrence of non-canopy dominating invasive species using remote sensing and environmental data in indigenous forests. For example Joshi et al. (2006) integrated Landsat ETM+ imagery with environmental data to map the probability of occurrence of *C. odorata* in south central Nepal forest. However, remote sensing data such as Landsat, SPOT or IKONOS consist of spectral bands whose signal tend to saturate in high canopy vegetation i.e. leaf area index (LAI) greater than three (Knipling 1970; Mutanga and Skidmore 2004). Subtropical forest gaps are usually characterised by a high LAI. The development of new generation high spatial resolution satellites such as RapidEye (5 meters) and WorldView-2 (2 meters) has opened opportunities for improved characterisation of vegetation in a high LAI environment(Ramoelo et al.

2012; Mutanga et al. 2012; Ozdemir and Karnieli 2011). The red edge band (700-725 nm) present in RapidEye and WorldView-2 has been shown to minimize the signal saturation problem common in traditional sensors, thereby improving the prediction of chlorophyll, nitrogen and vegetation biomass (Ramoelo et al. 2012; Mutanga et al. 2012; Cho et al. 2013). In remote sensing, the "red edge" is the region of abrupt change in the leaf reflectance between 680 and 780 nm due to the combined effect of strong chlorophyll absorption in the red and high reflectance in the near-infrared wavelength resulting from internal scattering in the spongy mesophyll (Horler et al. 1983). Could the presence of the red-edge band in WorldView-2 enhance the ability of predicting the occurrence of invasive *C. odorata* in forest canopy gaps?

Ancillary environmental data such as distance from roads/trails, distance from rivers, distance from forest edges that have been used to predict the distribution of species (Franklin 1995; Yang et al. 2006; Václavík and Meentemeyer 2009: Masocha and Skidmore 2011) can be easily generated from remote sensing imagery. The question is: Can the integration of ancillary environmental data with WorldView-2 data increase the accuracy of predicting the probability of occurrence of invasive *C. odorata* in forest canopy gaps? We used Dukuduku Coastal Forest of KwaZulu-Natal province, South Africa as a case study to test the afore-mentioned assumptions. Mapping the occurrence of the invasive species in forest canopy gaps rather than its actual cover is important since *C. odorata* occupies the canopy gaps mixed with other species, as opposed to growing understorey due its light, space and soil moisture requirement (Joshi et al. 2006).

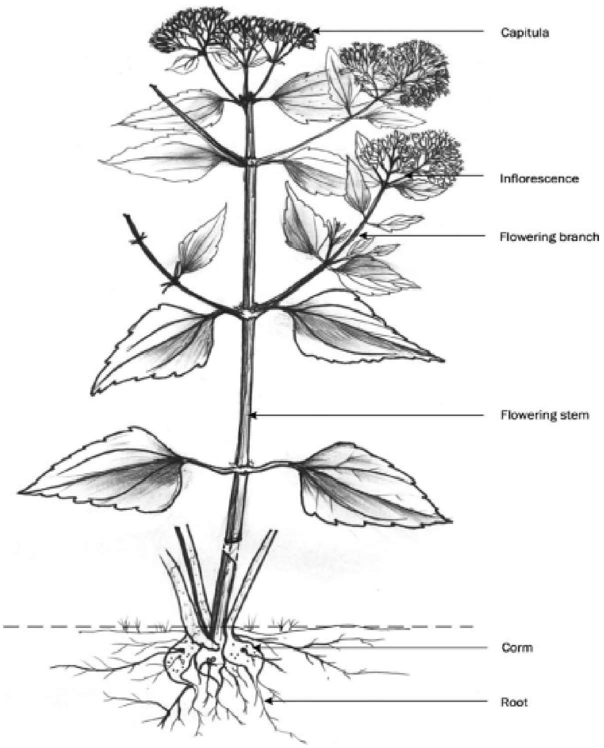


Fig. 1 Schematic representation of *Chromolaena odorata* morphology (Source: Joshi et al. 2006)

Study area

This study was undertaken in Dukuduku forest, the largest remaining patch of subtropical coastal forest in KwaZulu-Natal, South Africa (28°38'33"S and 32°31'67"E). (Fig. 2). The study area has an area of about 3172.43 hectares. On the western side, the forest is surrounded by the sugar plantation farms and the *Eucalyptus* plantations, while on the eastern side are villages that practise subsistence farming. The climate of KwaZulu-Natal is subtropical, with high summer precipitation and high temperatures of over 33°C (between September and April). Winters are generally cooler (below 8°C), with the annual sea surface temperature of 15°C. The area receives annual rainfall of about 1 600 mm (Luwum 2002).

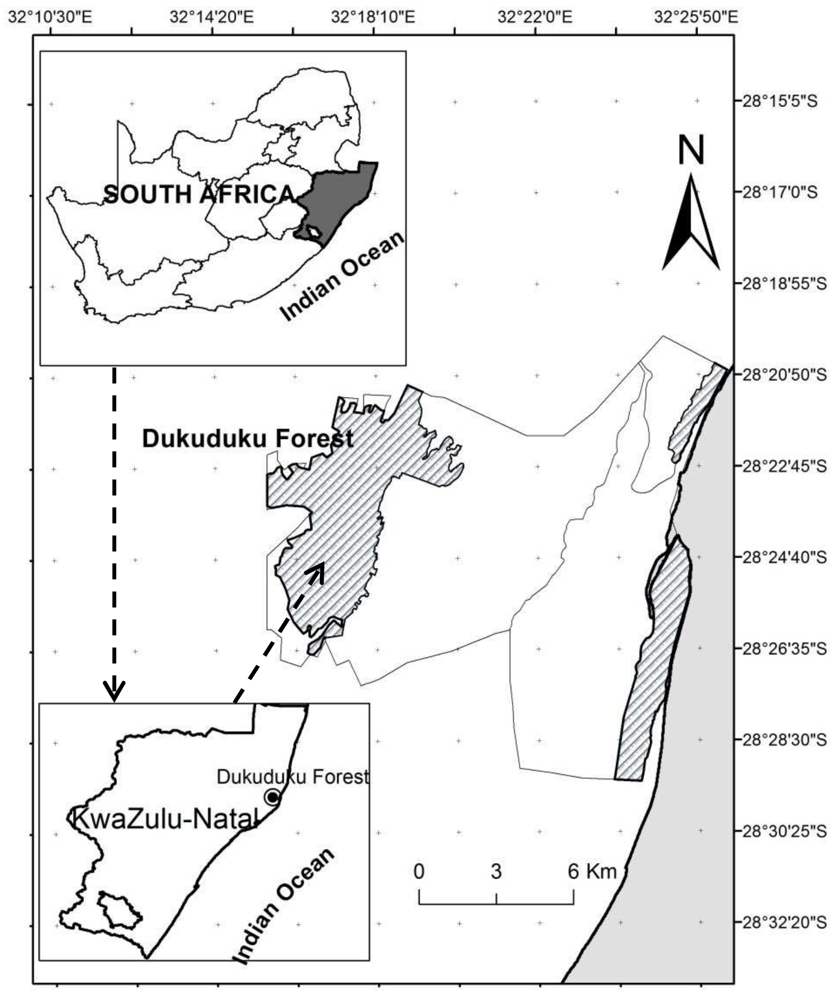


Fig. 2 The location of the Dukuduku forest in KwaZulu-Natal, South Africa.

Image acquisition and pre-processing

WorldView-2 image (acquired on 01 December 2010) with 8 spectral bands and at a 2 meter spatial resolution was used for the delineation of forest gaps. The delineated forest gaps (Fig. 3) were derived from WorldView-2 data using object-based image analysis (OBIA), with a 93.69% overall accuracy (Malahlela et al. in review). The image was geo-referenced to Universal Transverse Mercator (WGS 84), mosaicked and clipped to the study area. The imagery was geometrically corrected by the supplier (geolocation accuracy < 3.5m CE90, as specified by Digital Globe), and the atmospheric correction was done using the AtCOR 2/3 software distributed by ReSe® Applications. The atmospheric correction was based on the MODTRAN 5 module, which is a ‘narrow band model’ atmospheric radiative transfer code – with spectral range of between 0.2 to 100 µm (Berk et al. 1998). The atmospheric conditions specified in the AtCOR software for this image processing was the ‘tropical rural’ conditions due to the nature of the study area.

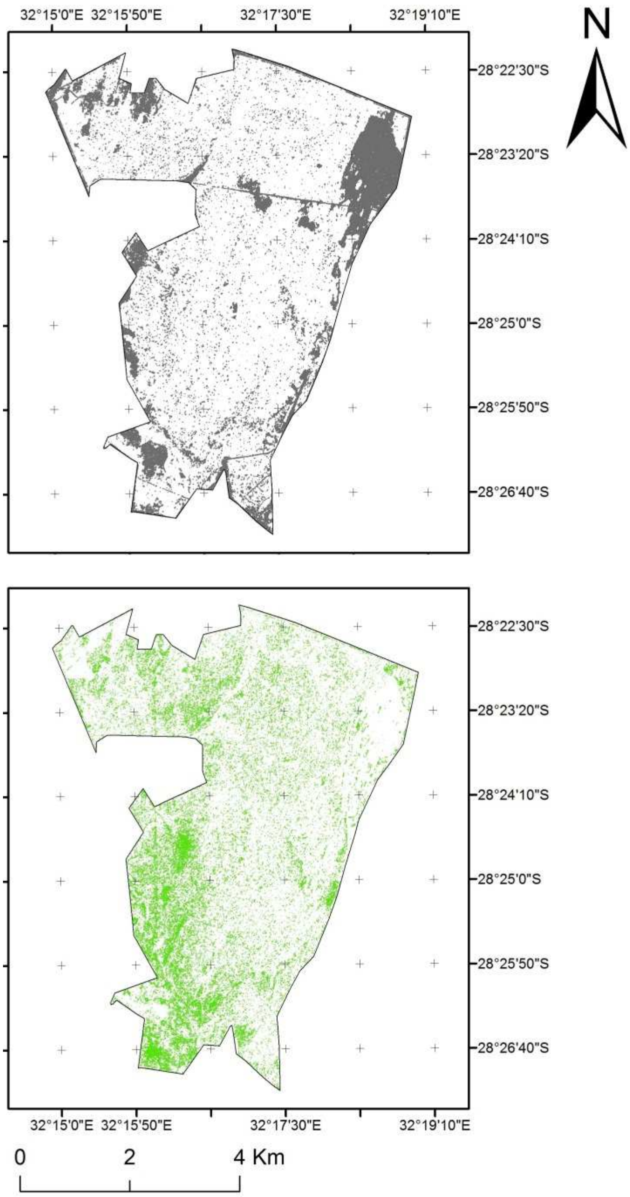


Fig. 3 Delineated forest gaps comprising of non-vegetated (upper) and vegetated (lower) gaps (Malahlela et al. in review).

Field data collection

Field data were collected on two occasions and in different seasons (July and October 2011) because of logistical constraints. The data collected included the location of forest canopy gaps (using Vista eTrex ™ GPS, with maximum spatial accuracy of 4 m), as well as the presence/absence of invasive species (*Chromolaena*) in forest gaps. The collection of the data followed a line transect method. A total of 13 line transects were visited across the forest, each with the minimum length of about 1 km. A simple random sampling technique was applied, where lines were randomly pre-selected to cover most parts of the forest. In total, 115 (n = 115) forest gaps were visited in the field. We used 2/3 (n =

77) of the data to construct a logistic regression model while 1/3 (n = 38) of the data was used to validate the model.

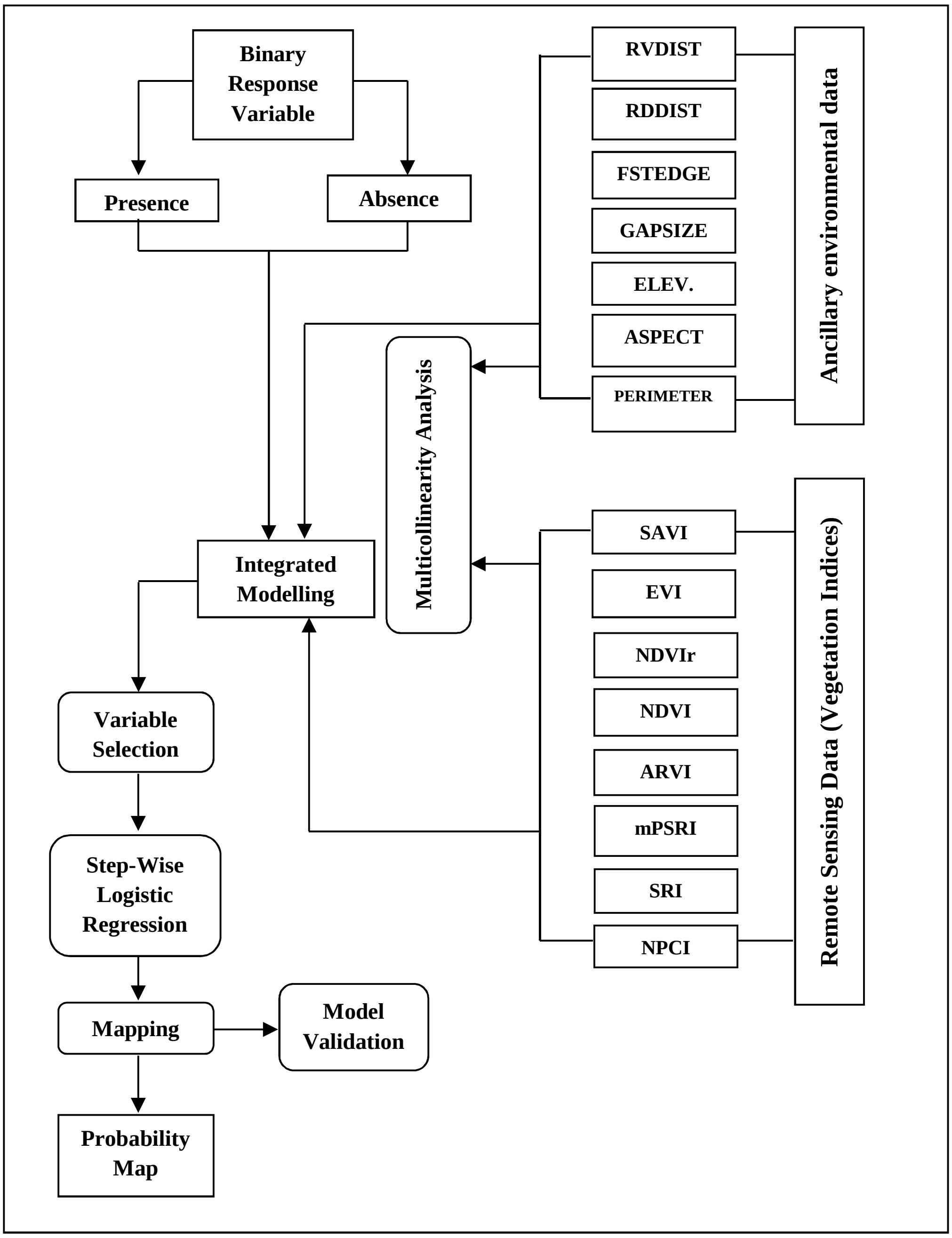


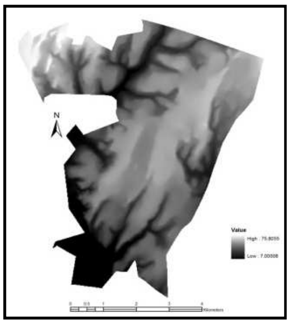
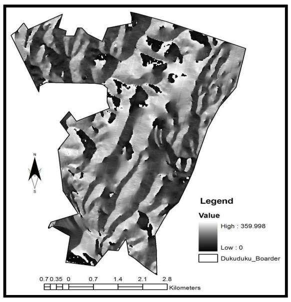
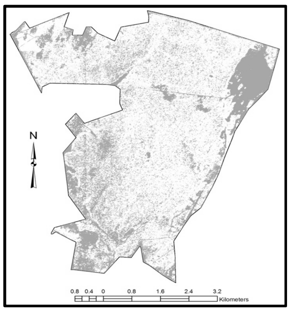
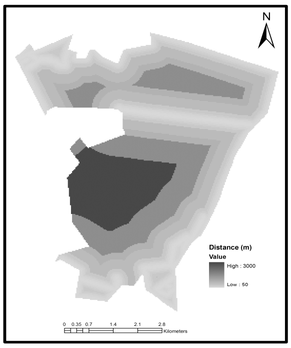
Fig. 4 Schematic representation of workflow followed during logistic regression modelling.

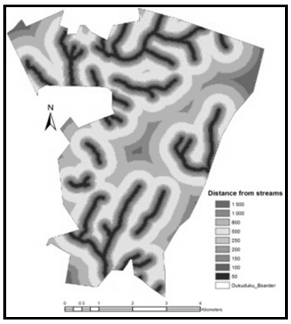
Data analysis

The relationship between *C. odorata* presence/absence, environmental data and remote sensing data was modelled in logistic regression as shown in Fig. 4. A logistic regression was conducted for 3 sets of variables, i.e. (a) environmental variables only (Table 1), (b) spectral variables only and (c) combined environmental and remote sensing variables in a stepwise logistic regression (Table 1).

Table 1

Ancillary environmental variables and WorldView-2 data variables used for modelling

Variable	Processing	Layer
Elevation (ELEV.)	The digital elevation model (DEM) layer was generated from the 15 meter contour lines of the study area Generation of the DEM is based on the algorithm ANUDEM, which calculates values on a regular grid of a discretised smooth surface fitted to a large number of irregularly spaced elevation data points, contour line data and stream line data (Hutchinson 1996).	
Aspect (ASPECT)	Aspect layer was generated from a 15 meter DEM with the help of Spatial Analyst tool in ArcGIS software	
Canopy gaps characteristics (size and perimeter) (GAPSIZE, PERIMETER)	Gap size and gap perimeter layers were generated from the WorldView-2 derived delineated forest canopy gaps. Forest canopy gaps were delineated in object-based image analysis using the modified plant senescence index (Merzlyak et al. 1999). This technique yielded the highest accuracy of 93.69% when compared to pixel-based classification, and therefore the results were used to derive forest gap size and perimeter.	
Distance from roads (RDDIST)	Distance from roads was generated from a road dataset in ArcGIS. The distances from roads were calculated from the forest gaps to the closest roads/paths. The forest's boundaries are mainly the main roads (national tar roads) and the gravel roads created for demarcating agricultural plantations. Distances were measured in meters (m)	

Distance from rivers (RVDIST)	The distance from river map was derived from river and streams dataset of the study area. The generation of this map was done in ArcGIS. The distance was calculated from the nearest stream to the forest gap to determine the influence of rivers to the distribution of invasive species. The distances were measured in meters (m).	
Vegetation indices	Two sets of vegetation indices were tested for the regression. These sets are (i) indices that are commonly derived from conventional sensors, and (ii) indices that can be derived from WorldView-2 imagery. A total of seven (7) vegetation indices were tested for this study (Table 8).	See table 2

Ancillary environmental variables

Ancillary environmental variables and WorldView-2 data variables were used for modelling. Ancillary environmental data were used based on their importance as factors driving the distribution of *C. odorata* as described in the introduction and availability. The ancillary environmental dataset included digital elevation model (DEM) layer, aspect layer (ASPECT)), distance from roads layer (RDDIST), distance from rivers layer (RVDIST), gap size layer (SIZE), distance from forest edges (FSTEDGE) and gap perimeter layer (PERIMETER) (Table 1).

Spectral data

Spectral dataset was processed as vegetation indices. We opted to use vegetation indices as individual bands on their own do not yield any significance relationships with species occurrence (Verstraete and Pinty 1996). A number of vegetation indices such as the Normalized Difference Vegetation Index (NDVI), the Red edge Normalized Difference Vegetation Index (NDVIr), the Normalized Green Vegetation Index (NDVIgr), the Soil-Adjusted Vegetation Index (SAVI), the modified Plant Senescence Reflectance Index (mPSRI), Simple Ratio Index (SRI), the Enhanced Vegetation Index (EVI), the Normalized Pigment Chlorophyll Index (NPCI), and the Atmospherically Resistant Vegetation Index (ARVI) were computed from WorldView-2 image These indices were treated individually as separate input variables for prediction. These indices fallen into either of the two sets, i.e. (i) indices that can be commonly derived from conventional sensors such as Landsat, and (ii) indices that can be derived from WorldView-2 imagery. A total of eleven (11) vegetation indices were tested for this study (Table 1and 2).

Table 2
Vegetation indices selected for the study

Index	Formula	Application	Reference
NDVI	$NDVI = \frac{\rho_{833}-\rho_{660}}{\rho_{833}+\rho_{660}}$	Traditional index used to monitor vegetation vigour, health, vegetation cover and biomass. Values range from -1(bare surfaces) to 1(green plants).	Jackson et al.(1983)
NDVIr	$NDVIr = \frac{\rho_{833}-\rho_{725}}{\rho_{833}+\rho_{725}}$	New WorldView-2 index that improves the detection of vegetation health, greenness, and biomass estimation. It is computed from the red edge band centered at 725nm, instead of a red band centered at 660nm as in NDVI. Values range from -1(bare surfaces) to 1(green vegetation).	Gitelson and Merzlyak (1994)
SRI (various)	$SRI = \frac{\rho_{660}}{\rho_{833}}$	Used for mapping vegetation health and condition. It is high in very healthy vegetation and low in stressed vegetation or non-vegetated areas.	Asrar (1989)
mPSRI	$mPSRI = \frac{\rho_{660}-\rho_{480}}{\rho_{725}}$	New WorldView-2 index used for detecting of leaf senescence, plant physiological stress and fruit ripening. Values range from -1(stressed canopy) to 1(less stressed canopy).	Merzlyak et al.(1999)
NDVIgr	$NDVIgr = \frac{\rho_{833}-\rho_{545}}{\rho_{833}+\rho_{545}}$	Traditional index that works similarly to NDVI and additionally measures the greenness of vegetation.	Gitelson et al.(1999)
EVI	$EVI = 2.5\left(\frac{\rho_{833}-\rho_{660}}{\rho_{833}+6\rho_{660}-7.5\rho_{480}+1}\right)$	The enhanced vegetation index (EVI) is an 'optimized' index designed to enhance the vegetation signal with improved sensitivity in high biomass regions and improved vegetation monitoring through a de-coupling of the canopy background signal and a reduction in atmospheric influences.	Huete et al.(1997)
ARVI	$ARVI = \frac{\rho_{833}-(2\rho_{660}-\rho_{480})}{\rho_{833}+(2\rho_{660}-\rho_{480})}$	An enhancement to the NDVI that is relatively resistant to atmospheric factors (for example, aerosol). It uses the reflectance in blue to correct the red reflectance for atmospheric scattering. It is most useful in regions of high atmospheric aerosol content, including tropical regions contaminated by soot from slash-and-burn agriculture.	Kaufman and Tanre (1996)
NPCI	$NPCI = \frac{\rho_{660}-\rho_{425}}{\rho_{660}+\rho_{425}}$	The normalized difference pigment chlorophyll index (NPCI) was developed especially for the detection of the chlorophyll content of crops.	Peñuelas et al. (1995)
SAVI	$SAVI = \left(\frac{\rho_{830}-\rho_{660}}{\rho_{830}+\rho_{660}+0.5}\right) * (1+0.5)$	Traditional index used also for vegetation monitoring, biomass and vegetation health. It improves on NDVI by compensating for soil-background.	Huete et al. (1997)

Model calibration

The dataset (n = 77) was randomly split into 2/3 (n = 38) for model calibration. The calibration dataset was used to train the model for invasive species occurrence using the R statistical software. A stepwise logistic regression model was used for all input variables. In stepwise logistic regression, an attempt is made to eliminate any insignificant variable from the model before adding a significant one to the model and to deal with multi-collinear variables. We used forward elimination procedure to select suitable variables in the final model (Manel et al. 1999). The choice of the forward stepwise logistic regression model was dictated by the binary nature of the response variable (presence/absence), its simplicity for embedding in GIS software (Yang et al. 2006) and its popularity amongst all other predictive models (Manel et al. 1999; Aspinall 2002). Logistic regression is given by the following equation:

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

where P is the probability of occurrence, x_n is the explanatory variable, β_n are the coefficient of x_n, β_0 is the intercept and e is the exponent function of the model. The final model goodness of fit was measured by the deviance (D^2) which is an analogy to a coefficient of determination (R^2) peculiar to logistic regression model (Rossiter and Loza 2012). The D^2 is obtained from the following equation:

$$D^2 = 1 - \left(\frac{residual\ deviance}{null\ deviance}\right) \quad (2)$$

Each variable removal or addition from or to a model is listed as a separate step in the model output. The model with the highest D^2 and lowest Aikaike's Information Criterion (AIC) was selected as the most ideal model because it has the best fit (Fox 2002).

Model validation

One-thirds of the data (n = 38) was used for validating the predictive model. The predicted probabilities (y), which ranged from values between 0 and 1, represented the increasing probability of *C. odorata* presence in forest canopy gaps. A range of thresholds was explored to determine the optimum threshold level for predicting *C. odorata* presence/absence (P/A) in forest gaps. The study by Manel et al. (1999) previously suggested a probability threshold value of 0.5 as the optimum threshold value for species prediction, although this value may not be ideal in all circumstances. For this study, we tested probability thresholds of 0.2 – 0.9. A 2 x 2 error matrix table (with rows indicating predicted cases and columns indicating observed cases) was plotted for a threshold value that yielded the highest mapping accuracy. The overall mapping accuracy is defined as the total number of the correctly predicted test cases to the total number of test samples, and is presented as a percentage (Fielding and Bell 1997). The table compares the predicted values (from an optimum threshold value) with the observed field data of *C. odorata* distribution.

The area under the ROC (AUC) has been used in several studies in order to understand the robustness of the model for a binary classifier (Egan 1975; Swets et al. 2000; Fawcett 2006; MedCal 2014). The AUC value of 0.5 indicates that the model accuracy is equal to the random prediction, while the value of 1.0 shows the perfect model fit (Baldwin 2009). In essence, the AUC has a quantitative measure on a 0.0 to 1.0 scale, with the following grading levels:

- < 0.6 indicates a poor model

- 0.6 – 0.7 indicates a pass model
- 0.7 – 0.8 indicates a good model
- >0.9 indicates an excellent model

Furthermore, the sensitivity and specificity analysis was performed across the probability range from 0.2 – 0.9. For binary error matrix, sensitivity is defined as the proportion of correctly classified presence to the total number of presences in the test samples. On the other hand, specificity is the proportion of correctly predicted absence to the total number of absence in test samples (Fielding and Bell 1997).

Results

Logistic regression

The combined environmental and spectral data model explained 71 % of the variance in the *C. odorata* presence/absence data ($D^2 = 0.71$, $p < 0.05$), which was the highest when compared to the spectra data only ($D^2 = 0.30$, $p < 0.05$) and ancillary environmental data only ($D^2 = 0.52$, $p < 0.05$) models. From the integrated stepwise logistic model, two of the environmental variables (distance from rivers and distance from roads) were significant at $p < 0.05$. These environmental variables have shown negative relationship to the presence/absence of *C. odorata* in the forest gaps, at $p < 0.05$.

Table 3
The results of three logistic regression models and their significant (shown by * and º) and non-significant variables.

Data Source	Predictor	Estimate	Std. Error	*z* value	*p* value
Environmental Variables ($D^2 = 0.52$)	(Intercept)	5.9560	2.2394	2.2660	0.007 **
	Distance from rivers	- 0.0059	0.0023	- 2.601	0.009 **
	Distance from roads	- 0.0033	0.0014	- 2.400	0.016 *
	Elevation	- 0.0230	0.0334	- 0.689	0.490
	Aspect	0.0029	0.0051	0.582	0.561
	Gap Size	0.0275	0.0200	1.372	0.169
	Distance from edges	- 0.0013	0.0013	- 0.982	0.326
	Gap perimeter	- 0.0509	0.0363	- 1.401	0.161
WorldView-2 Variables ($D^2 = 0.30$)	(Intercept)	- 273.080	273.274	- 0.999	0.317
	NDVIgr	14.506	14.632	0.991	0.322
	SRI_r	803.929	380.604	2.112	0.035 *
	SRI_IR	- 1.900	1.021	- 1.861	0.063 °
	SRI_re	- 281.530	143.338	- 1.964	0.050 *
	NDVIr	- 361.554	190.469	- 1.898	0.058 °
	mPSRI	- 347.555	163.974	- 2.120	0.034 *
	SAVI	1861.780	1066.201	1.746	0.081 °
	NDVI	- 1907.697	1460.870	- 1.306	0.192
	ARVI	- 322.005	192.147	- 1.676	0.093 °
	NPCI	10.572	21.089	0.501	0.616
	EVI	- 0.986	4.297	- 0.229	0.819
Combined Model ($D^2 = 0.57$)	(Intercept)	114.600	62.26	1.840	0.066 °
	Distance from rivers	- 0.0200	0.0100	- 2.121	0.033 *
	Distance from roads	- 0.0100	0.0000	- 2.225	0.026 *
	NDVIr	- 38.110	24.770	- 1.538	0.123
	SAVI	- 76.920	42.190	- 1.823	0.068 °
	mPSRI	- 291.80	167.70	- 1.740	0.081 °

NDVIgr	68.910	51.750	1.823	0.183
EVI	- 19.000	10.780	- 1.763	0.078 ˚
Elevation	- 0.0900	0.0500	- 1.922	0.050 ˚

Significance codes: (˚), 0.1 (*), 0.05 (**), 0.01 (***), 0,001

The elevation was the only environmental variable with a significant negative relationship to the presence/absence of invasive species ($p < 0.05$). Additionally, from this model, three of the spectral data variables (mPSRI, SAVI and EVI) were significant at *alpha* < 0.1. All of the vegetation indices have shown negative correlation to the invasive *C. odorata* presence/absence in forest gaps. On the other hand, the NDVIgr showed a positive relationship to the presence/absence data of *C. odorata* in forest gaps (Table 3). The results from environmental data-only model have shown that the intercept ($p < 0.01$), distance from roads ($p < 0.01$), and distance from roads/paths ($p < 0.05$) were significantly correlated to species presence/absence. Among the environmental data only model, the distance from rivers was the most significant positive variable at $p < 0.01$. Amongst WorldView-2 data only model, six of the variables were significant, with red Simple Ratio Index (SRI_r, computed from red/NIR1), the red edge Simple Ratio Index (SRI_re, computed from red edge/NIR1) and mPSRI were significantly correlated to the species presence/absence in the forest gaps at $p < 0.05$. Figure 5 shows the predicted probability of occurrence of invasive species across the delineated forest gaps.

Overall, the integrated step-wise logistic regression model has shown an improvement of over 36% compared to environmental variables only.

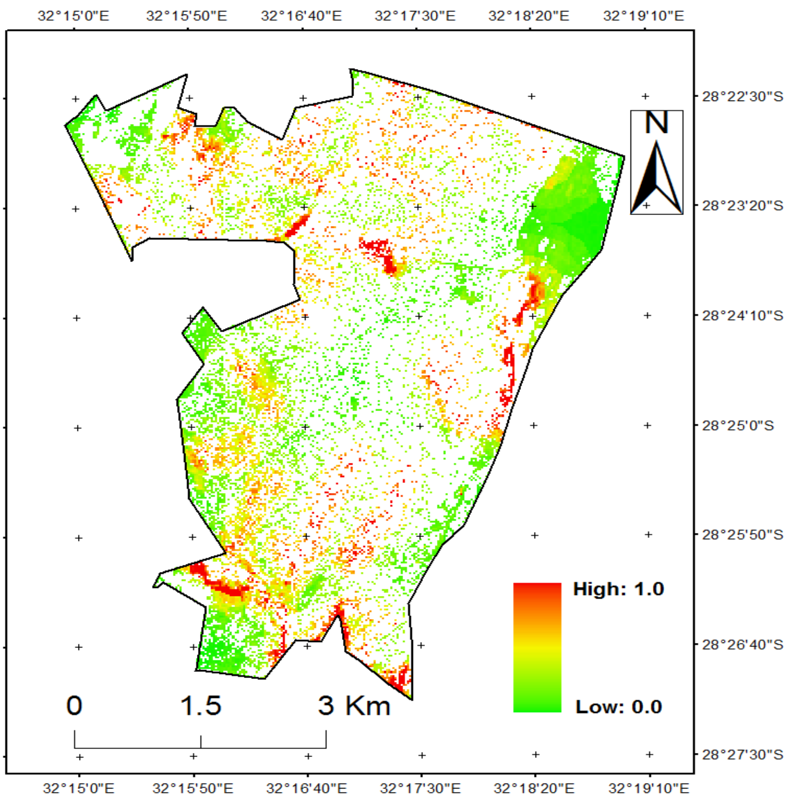


Fig. 5 A predicted probability map indicating the occurrence of *Chromolaena odorata* in delineated forest gaps at the Dukuduku forest.

Model validation

The predictive model (binary outcome value range of 0.0 – 1.0) was validated using probability threshold values as shown in table 4. The highest prediction accuracies were obtained at threshold range between 0.2 and 0.3 (both at 87%). The highest sensitivity rates (defined as the proportion of correctly classified presence to the total number of presences in the test samples) were observed at similar threshold range (87%). The highest specificity rates (specificity is the proportion of correctly predicted absence to the total number of absence in test samples) were obtained at the threshold values of 0.8 and 0.9 (both at 86%) (Fielding and Bell 1997). The 2 x 2 error matrix table for a probability threshold of 0.2 and 0.3 is shown in table 5.

Table 4
Statistics for evaluating model performance across probability threshold values

	Probability Threshold							
	0.2	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
Prediction Accuracy (%)	0.87	0.87	0.84	0.84	0.84	0.82	0.82	0.79
Sensitivity (%)	0.87	0.87	0.84	0.84	0.84	0.81	0.81	0.77
Specificity (%)	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86

Table 5
Predicted outcomes (*y*) from logistic regression on *Chromolaena odorata* vs. the observed field data at probability threshold of 0.2 and 0.3 (ρ = 0.3)

		Predicted Occurrences		
		Presence	Absence	Total
Observed Occurrences	Presence	27	4	31
	Absence	1	6	7
	Total	28	10	38

The robustness of the model (curve) was also measured by the Area Under the Curve of receiver operating characteristic (ROC) curve (Fig. 6). The validation dataset yielded an AUC of 0.89 at p = 0.001, which shows that the model used for prediction was significant. The diagonal line in the model represents the strategy of randomly guessing a class (Fawcett 2006). If the curve bends towards or below the diagonal line (decreasing sensitivity) this indicates a poor model. A good model is the one whose curve bends towards the north-western direction of the plot (Fawcett 2006). Our ROC shows that our curve bends towards the north-western direction, and hence an AUC of 0.89.

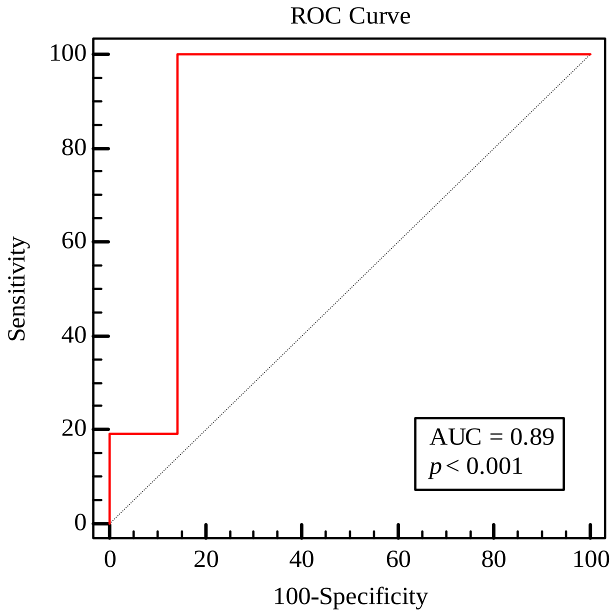


Fig. 6 The ROC curve derived from validation dataset (n = 38) across different probability thresholds. The dotted line indicates a line of no-discrimination (random guess) while the red line indicates sensitivity and specificity of model across different threshold values.

Discussion

The findings of the study suggest that there is a relationship between WorldView-2 spectral bands and the presence/absence of invasive *C. odorata* (Table 3). The model that combined the environmental variables and spectral variables yielded the highest prediction accuracy, which underscores the importance of such variables in the prediction of *C. odorata* in forest gaps (Ozdemir and Karnieli 2011) (Fig. 7). This integration is necessary since on their own, the ancillary environmental variables or WorldView-2 data do not yield prediction accuracies greater than 60%. The general trend depicted by the predictive model shows that the probability of occurrence of *C. odorata* tends to increase in forest gaps that are less vegetated than those that are densely vegetated. This trend is supported by observing the predictive model's spectral estimates of the NDVIr, mPSRI, SAVI and EVI, most of which are significantly correlated with the presence and absence data of *C. odorata*. A negative estimate of red edge band (used to compute NDVIr) means that the presence of *C. odorata* decreases with increasing density of vegetation, as an increase in reflectance at this spectral region (705 – 745 nm) is associated with increases in vegetation densities or biomass (Knipling 1970). Similar findings were achieved by Joshi et al. (2006) who observed that *C. odorata* does not thrive in densely vegetated forest areas. This characteristic is very important to this invader, in that it satisfies the light requirement of the species in question. These findings show that the red edge is significant in mapping the probability of occurrence of *C. odorata*.

Previous studies have highlighted that the plant senescence reflectance index (PSRI, from which mPSRI was derived) is sensitive for detecting senescing leaves and to detect physiological stress at different developmental stages of plants (Merzlyak et al 1999; Peñuelas et al 1994; Hatfield and Prueger 2010). In the same light the mPSRI (which uses blue and red edge band, instead of green and NIR) indicated that increase in plant stress in vegetated forest gaps increases the probability of *C. odorata* presence. This is true when observing the negative estimate of mPSRI (-291.80, $p < 0.08$)

since a negative estimate indicates the stressed vegetated area, while the positive estimate indicates less stressed vegetated areas (Merzlyak et al 1999). The enhanced vegetation index (EVI) has long been used to assess vegetation biomass in different biomes, and it is already established that increase in EVI values is associated with increase in the density of vegetation (Huete et al 1997). Conversely, our study has found that decreases in EVI values is associated with the probability of invasive species presence in forest gaps, which is an indication that less vegetated forest gaps are prone to invasion than their vegetated counterparts. These findings are especially true considering the negative correlation between invasive species distribution and SAVI, which is known to improve on NDVI by compensating for soil background (Huete et al. 1997).

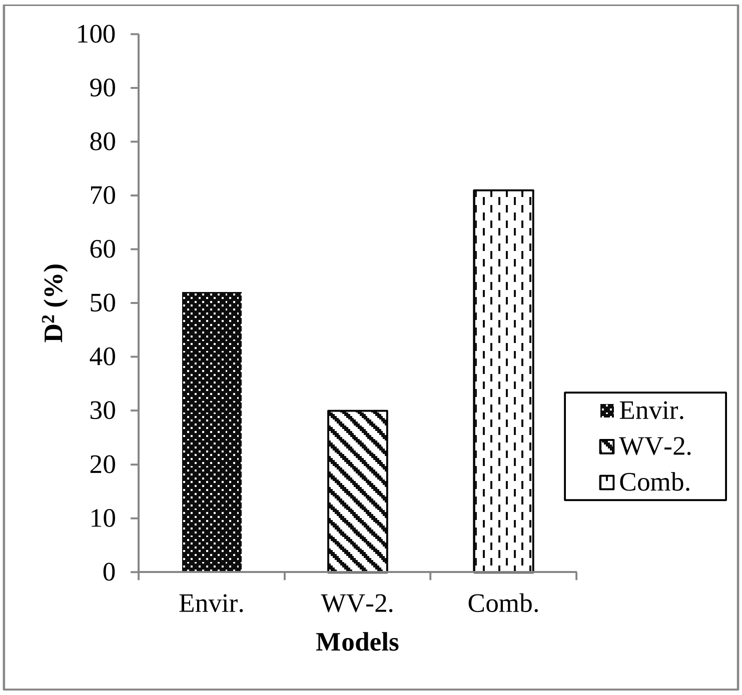


Fig. 7 Model accuracy comparison amongst environmental data-only model (Envir), WorldView-2 data-only (WV-2) and a combined environmental and WorldView-2 data model (Comb).

The environmental variables such as distance from rivers/streams and aspect were found to be negatively significant in determining the presence and absence of *C. odorata* occurrence. The implication is that as one advances closer to the streams, the probability of finding *C. odorata* increases, and this is in line with the findings by Joshi et al. (2006) and Van Gils et al. (2006) who found areas closer to roads and edges to be more likely invaded by *C. odorata*. The tap root system of this species gives it a competitive advantage over water and ensures its stability in invaded habitats. The results also showed that the probability of invasive species occurrence increases with the decrease in elevation. This is true since in troposphere an increase in elevation results in decrease in temperature, and thereby inhibiting the occurrence of plant species that are adapted to warmer conditions. Additionally, one is more likely to find invasive species in north facing slopes, than the south facing slopes, as shown in Table 3. This is due to the fact that north-facing slopes in the southern hemisphere are warmer than south-facing slopes (Adams 2010). The species triumphs in areas that are open, with appropriate light and temperature ranges of between 20 - 37˚C, and hence the increase in slope direction (towards the north) increases probability of finding *C. odorata* (Gareeb 2007). Our study also highlighted that invasive *C. odorata* prefers forest gaps that are closer to the streams due to their competitive nature for water and essential mineral resources. This trend is in line with the findings by Goodall and Zacharias (2002) who observed that this invasive species prefers forest margins and is often found along rivers or streams. Roads and forest edges serve as the alleys/corridors through which the seeds of this species are dispersed, especially considering the fact that they require wind as a dispersal mechanism. On the whole, the pattern of the probability of occurrence (fig. 5) indicate that *C. odorata* is less likely to occur in pristine forest than in the areas that are open, such as closer to the roads or at the edges of forest (Joshi et al. 2006).

Management of invasive species

The management of invasive species such as *C. odorata* has been debated in different countries, globally. For example Herren-Gemill (1991) described the need to control *C. odorata* due to its high frequency of occurrence in invaded fallow sites in West Africa. From the conservation point of view, management and control priority should be focused on the species' habitat and future distribution of species (Rowe 1992), and not solely on the degradation levels caused by this species (Goodall and Erasmus 1996). Mapping the probability of occurrence of invasive species in its potential habitats (forest gaps) is crucial to the management geared towards eliminating such species. The output maps serve as a guideline to fieldworkers and forest managers for the identification of probable habitats of *C. odorata* and for man-power recruitment.

Fig. 8 Probability maps are crucial to the management of invasive species. In these picture frames the researcher (Primary author of this paper) presents the probability maps to the field workers who were tasked to eradicate invasive *Chromolaena odorata* in across the forest.

In South Africa, for example, there are projects that are aimed at eradicating invasive *C. odorata* in the coastal forests, such as the Dukuduku forest (Fig.8), where field workers are assigned to walk randomly through the forest to eradicate visible invasive species. The modeling of invasive species probability of occurrence could potentially assist in eliminating the random search of invasive species by providing key indications of areas of high probability of occurrence. Additional environmental data variables such as mean annual temperature, precipitation, soil data (pH, texture, moisture), could potentially improve the prediction power of the model. The application of predictive models such as Maxent has also shown to increase the prediction accuracy of species presence/absence data but has not been used for the study (Kumar and Stohlgren, 2009).

3.6 Conclusion

The additional bands present in WorldView-2 bands increases the capability of the sensor in the mapping of the probability *C. odorata* presence/absence in subtropical forest canopy gaps when compared to ancillary environmental variables. The improved accuracies are derivable when using WorldView-2 data products such as vegetation indices. The environmental data-only model explained about 52% of the presence/absence of invasive *C.odorata* presence or absence in forest gaps. The final combined model of WorldView-2 spectral data and ancillary environmental data increased predictive model accuracy to 71% (D^2 = 0.71; WorldView-2 data added) from 52 % (D^2 = 0.52; environmental data only), which emphasizes the advantage of integrating WorldView-2 with the ancillary environmental data for invasive species mapping. From the selected model, all other vegetation indices have shown the expected pattern of the distribution of *C. odorata* in forest gaps, except the green normalized difference vegetation index (NDVIgr), which has shown to be insignificantly positively related to invasive species in forest gaps. Although this variable contributed to the model, its implication to invasive species occurrence is subject to further investigation. Exploring the indices centered on new bands of WV-2 such as coastal band, yellow band and near-infrared-2 could potentially increase the accuracy of prediction, rather than red edge-centric analysis.

Acknowledgments

We wish to thank the Council for Scientific and Industrial Research's Natural Resources and Environment (CSIR-NRE) unit, of South Africa, for its financial assistance through its research grants. Many thanks also to the Department of Agriculture, Forestry and Fisheries for the support on the study area. Authors like to thank the Department of Science and Technology of South Africa for their financial support. Many thanks to Dr. R Mathieu for his leadership role at Earth Observation Group of NRE and Dr. Ramoelo for assisting with field work and insights on this paper. We appreciate the statistics tutorship assistance offered by Dr. Pravesh Debba and Nontembeko Dudeni-Tlhone of CSIR's Built Environment (CSIR-BE). Authors wish to thank Sibuyiselo Gumede of Khula village who assisted in field data collection. To Tino, my dearest son.

References

Adams J (2010) Vegetation – Climate Interaction. "How Plants Make the Global Environment". Second Edition, Springer. ISBN 978-3-642-00880-1.

Adams J (2010) Vegetation – Climate Interaction. "How Plants Make the Global Environment". Second Edition, Springer. ISBN 978-3-642-00880-1.

Asner GP, Knapp DE, Kennedy-Bowdoin T, Jones MO, Martin RE, Boardman J, Hughes RF (2008) Invasive species detection in Hawaiian rainforests using airborne imaging spectroscopy and LiDAR. Rem Sens of Envir 112: 1942-1955.

Aspinall R (2002) The use of logistic regression for validation of maps of the spatial distribution of vegetation species derived from high spatial resolution hyperspectral remotely sensed data. Ecol Model 157: 301-312.

Asrar G (1989) Theory and applications of optical remote sensing. John Wiley & Sons, Inc. New York – Chichester – Brisbane – Toronto – Singapore: 734.

Awanyo L, Attuah EM, McCarron M (2011) Rehabilitation of forest-savannas in Ghana: the impacts of land use, shade, and invasive species on tree recruitment. Appl Geogr 31: 181-190.

Baldwin RA (2009) Use of Maximum Entropy Modelling in Wildlife Research. Entr 11: 854-866.

Balée W (1989) The culture of Amazonian forests. In. D.A.Posey and W. Balée, eds., Resource Management in Amazonia: Indigenous and Folk Strategies. Adv in Econ Bot 7: 1-21. Bronx: New York Botanical Garden.

Berk A, Bernstein L, Anderson G, Acharya P, Robertson D, Chetwynd J, Adler-Golden S (1998) MODTRAN cloud and multiple scattering upgrade with application to AVIRIS. Rem Sens of Envir 65: 367-375.

Bousquet P, Peylin P, Ciais P, Le Quéré C, Friedlingstein P, Tans PP (2000) Regional Changes in Carbon Dioxide Fluxes of land and Oceans Since 1980. Sci 290: 1342-1346.

Brokaw NVL (1982) The definition of Treefall gap and its effect on measures of forest dynamics. Biotrop 14: 158-160.

Buckland ST, Elston DA, Beaney SJ (1996) Predicting distributional change, with application to bird distributions in northeast Scotland. Glob Ecol and Biogeo Lett 5: 66-84.

Chen Q (2009) Improvement of the Edge-based Morphological (EM) method for lidar data filtering. Int J of Rem Sens 30: 1069-1074.

Cho MA, Ramoelo A, Debba P, Mutanga O, Mathieu R, Van Deventer H, Ndlovu N (2013) Assessing the effects of subtropical forest fragmentation on leaf nitrogen distribution using remote sensing data. Landsc Ecol 28: 1479-1491.

Cunningham AB (2001) Applied ethnobotany: people, wild plant use and conservation. Earthscan, London.

de Rouw A (1991) The invasion of *Chromolaena odorata* (L.) King and Robinson (ex *Eupatorium odoratum*), and competition with the native flora, in a rain forest zone, south-west Cote d'Ivoire. J of Biogeo 18: 13 – 23.

Edwards TC, Cutler DR, Beard KH, Gibson J (2007) Predicting invasive plant species occurrences in national parks: a process for prioritizing prevention. Final Project Report No. 2007-1, USGS Utah Cooperative Fish and Wildlife Research Unit, Utah State University, Logan, UT 84322-5290 USA.

Egan JP (1975) Signal detection theory and ROC analysis, Series in Cognition and Perception. Academic Press, New York.

Fawcett T (2006) An introduction to ROC analysis. Patt Recog Lett 27: 861-874.

Fielding AH, Bell JF (1997) Review methods for the assessment of prediction errors in conservation presence/absence models. Envir Conserv 24: 38-49.

Fourcade, H.G., 1889, Report on the Natal forests, 197 pp. W. Watson, Pietermaritzburg.

Fox J (2002) Linear mixed models - An R and S-PLUS Companion to Applied Regression. R Development Core Foundation.

Franklin J (1995) Predictive vegetation mapping: geographic modeling of biospatial patterns in relation to environmental gradients. Prog in Phys Geog 19: 474-499.

Gareeb M (2007) Investigating in the potted *Chromolaena odorata* (L.) R.M. King and H. Robinson (*Asteraceae*). Thesis: School of Biological and Conservation Sciences, University of KwaZulu-Natal, Durban South Africa, 3-48.

Geldenhuys, C. J., 1989, Biogeography of the mixed evergreen forests of southern Africa. Ecosystems Programmes Occasional Report no. 45. FRD, Pretoria.

Gitelson A, Merzlyak MN (1994) Spectral reflectance changes associated with autumn senescence of *Aesculus hippocastanum* L. and *Acer platanoides* L. leaves. Spectral features and relation to chlorophyll estimation. J Plant Physiol 143:286-292.

Gitelson AA, Stark R, Kaufman YZ, Rundquist D (1999) A Technique for remote estimation of percent vegetation fraction, In Proc. the International Symposium on Spectral Sensing Research, Systems and Sensors for the New Millennium, *Int. Soc for Photogry and Rem Sens*, Oct. 31-Nov. 4, 1999, Las Vegas NV,191-201.

Goodall JM, Erasmus DJ (1996) Review of the status and integrated control of the invasive weed, *Chromolaena odorata*, in South Africa. Agric, Ecos and Envir 56: 151-164.

Goodall JM, Zacharias P (2002) Managing *Chromolaena odorata* in subtropical grasslands in KwaZulu-Natal, South Africa. In: Zachariades, C., Muniappan, R., Strathie, L (eds). *Proceedings of the 5th international workshop on biological control and management of Chromolaena odorata*. ARC-PPRI, Pretoria, South Africa: 120-127.

Gu W, Swihart RK (2004) Absent or undetected? Effects of non-detection of species occurrence on wildlife – habitat models. Biol Conserv 116: 195-203.

Harding GB, Bate GC (1991) The occurrence of invasive *Prosopis* species in the northe-western Cape, South Africa. S Afr J of Sci 87: 188-192.

Hatfield JL, Prueger JH (2010) Value of suing different vegetative indices to quantify agricultural crop characteristics at different growth stages under varying management practices. Rem Sens 2: 562-578.

Herren-Gemmill B (1991) The ecological role of the exotic *Asteraceous Chromolaena odorata* in the bush fallow farming system of West Africa. Biotrop. Special Publication 44: 11-22.

Hirzel AH, Guisan A (2002) Which is the optimal sampling strategy for habitat suitability modelling? Ecol Mod 157: 33-41.

Horler DNH, Dockray M, Barber J (1983) The red edge of plant leaf reflectance. Int J of Rem Sens 4: 273-288.

Huete AR, Liu HG, Batchily K, Leeuwen, WJ (1997) A Comparison of Vegetation Indices over a Global Set of TM Images for EOS-MODIS. Rem Sens of Envir 59:440-451.

Hutchinson M F (1996) A locally adaptive approach to the interpolation of digital elevation models. In Proceedings, Third International Conference/Workshop on Integrating GIS and Environmental Modeling, Santa Fe, NM, January 21-26, 1996. Santa Barbara, CA: National Center for Geographic Information and Analysis.

Jackson RD, Pinter Jr PJ, Reginato RJ (1983) Discrimination of growth and water stress in wheat by various vegetation indices through clear and turbit atmospheres. Rem Sens of Envir 13: 187-208.

Joshi C, de Leeuw J, Skidmore AK, van Andel J, Lekhak HD, van Duren IC (2006) Indirect remote sensing of a cryptic forest understorey invasive species. For Ecol Mngt 225: 245–256.

Kaufman YJ, Tanré D (1996) Direct and Indirect Methods for Correcting the Aerosol Effect on Remote Sensing. Rem Sens of Envir 55: 65-79.

Knipling EB (1970) Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation. Rem Sens of Envir 1: 155-159.

Kumar S, Stohlgren TJ (2009) Maxent modeling for predicting suitable habitat for threated and endangered tree *Canacomyrica monticola* in New Caledonia. J of Ecol and Nat Envir 1: 95-98.

Kupfer, JA, Runkle JR (1996) Early gap successional pathways in a *Fagus-Acer* forest preserve: patterns and determinants. J. Veg. Sci. 7: 247-256.

Laba M, Smith S, Richmond ME (2004) Purple loosestrife research and mapping for the Hudson River Valley study area. Final Report, New York Cooperative Fish and Wildlife Research Unit, Department of Natural Resources, Cornell University, Ithaca, NY.

Le Maitre DC (2002) Invasive alien trees and water resources in South Africa: case studies of the costs and benefits of management. For Ecol Mngt 160: 143-159.

Luwum P (2002) Control of Invasive *Chromolaena odorata*: An evaluation in some land use types in KwaZulu Natal, South Africa.

Malahlela, OE, Cho MA, Mutanga O (in review) Mapping canopy gaps in an indigenous subtropical coastal forest using high resolution WorldView-2 data. Int J of Rem Sensing.

Manel S, Dias JM, Ormerod SJ (1999) Comparing discriminant analysis, neural networks and logistic regression for predicting species' distributions: a case study with a Himalayan river bird. Ecol Model 120: 337-347.

Margules CR, Pressey RL (2000) Systematic Conservation Planning. Nat 405: 243-253.

Masocha M, Skidmore AK (2011) Integrating common classifiers with a GIS expert system to increase the accuracy of invasive species mapping. Int J Appl Earth and Observ Geoinf 13: 487-494.

MedCal (2014) Statistics for biomedical research software. Version 13.0.

Merzlyak JR, Gitelson AA, Chivkunova OB, Rakitin VY (1999) Non-destructive Optical Detection of Pigment Changes During Leaf Senescence and Fruit Ripening. Phys Plant 106: 135-141.

Mutanga O, Adam E, Cho MA (2012) High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm. Int J of Appl Earth Observ and Geoinf 18: 399-406.

Mutanga O, Skidmore AK (2004) Narrow band vegetation indices solve the saturation problem in biomass estimation. Int J of Rem Sens 25: 3999-4014.

Ozdemir I, Karnieli A (2011) Predicting forest structural parameters using the image texture derived from WorldView-2 multispectral imagery in a dryland forest, Israel. Int J of Appl Earth Observ and Geoinf 13: 701-710.

Peñuelas J, Baret F, Filella I (1995) Semi-empirical indices to assess carotenoids/chlorophyll a ratio from leaf spectral reflectance. Photosyn 31: 221-230.

Peñuelas J, Gamon JA, Fredeen AL, Merino J, Field CB (1994) Reflectance indices associated with physiological changes in nitrogen- and water-limited sunflower leaves. Rem Sens of Envir 48: 135 - 146.

Ramoelo A, Skidmore AK, Cho MA, Schlerf M, Mathieu R, Heitkonig IMA (2012) Regional estimation of savanna grass nitrogen using the red-edge band of the spaceborne RapidEye sensor. Int J of Appl Earth Observ and Geoinf 19: 151-162.

Reyers B (2004) Incorporating anthropogenic threats into evaluations of regional biodiversity and prioritisation of conservation areas in the Limpopo Province, South Africa. Biol Conserv 118: 521-531.

Rossiter DG, Loza A (2012) Analyzing land cover change with logistic regression in R. University of Twente, Faculty of Geo-Information Science & Earth Observation (ITC), Enschede (NL).

Rowe S (1992) The ecosystem approach to forestland management. Forest Chron 680: 222-224.

Sahid IB, Sugau JB (1993) Allelopathic effect of *Lantana camara* and *Chromolaena odorata* on selected crops. Weed Sci 41: 303 – 308.

Scott JM, Heglund PJ, Morrison ML, Haufler JB, Raphael MG, Wall WA, Samson FB (Eds.) (2002) Predicting Species Occurrences: Issues of Accuracy and Scale. Island Press, Washington, DC.

Skole DL, Tucker CJ (1993) Tropical deforestation and habitat fragmentation in the Amazon: satellite data from 1978 to 1988. Sci 260: 1905-1910.

Sohngen B, Alig R (2000) Mitigation, Adaptation, and Climate Change: results from Recent Research on U.S. Timber markets. Envir Sci Pol 3:235-248.

Suarez AV, Bolger DT, Case TJ (1998) Effects of fragmentation and invasion on native ant communities in coastal southern California. Ecol 79: 2041-2056.

Swets JA, Dawes RM, Monahan J (2000) Better decisions through science. Scient American 283: 82-87.

Turner W, Spector S, Gardiner N, Fladeland M, Sterliug E (2003) Remote sensing of biodiversity science and conservation. Trends in Ecol & Evol 18: 306-314.

Underwood E, Ustin S, DiPietro D (2003) Mapping non-native plants using hyperspectral imagery. Rem Sens of Envir 86: 150-161.

Underwood E, Ustin S, DiPietro D (2003) Mapping non-native plants using hyperspectral imagery. Rem Sens of Envir 86: 150-161.

Ustin SL, Gitelson AA, Jacquemoud S, Schaepman M, Asner GP, Gamon JA (2009) Retrieval of foliar information about plant pigment systems from high resolution spectroscopy. Rem Sens of Envir 113: (Supplement 1): S67-S77.

Václavík T, Meentemeyer R (2009) Invasive species distribution modeling (iSDM): Are absence data and dispersal constraints needed to predict actual distributions? Ecol Model 220: 3248-3258.

Van Gils H, Mwanangi M, Rugege D (2006) Invasion of an alien shrub across four land management regimes, west of St Lucia, South Africa. S Afr J of Sci 102: February.

Verstraete MM, Pinty B (1996) Designing optimal spectral indexes for remote sensing applications. IEEE Trans on Geosci and Rem Sens 34: 1254-1265.

Whitmore TC (1989) Canopy gaps and the two major groups of forest trees. Special feature – Treefall gaps and forest dynamics. Ecol 70: 536–538.

Witkowski ETF, Wilson M (2001) Changes in density, biomass, seed banks and seed production of the alien invasive plant *Chromolaena odorata*, along a 15 year chronosequence. Plant Ecol 152: 13-27.

World Wildlife Fund (2013). www.wwf.org . Accessed on the 12 December 2013.

Yang X, Skidmore AK, Melick DR, Shou Z, Xu J (2006) Mapping non-wood forest product (matsutake mushrooms) using logistic regression and GIS expert system. Ecol Model 198: 208-218.