

Assessing the Utility of the SPOT 6 Sensor in Detecting and Mapping *Lantana camara* for a Community Clearing Project in KwaZulu-Natal, South Africa

Zakariyyaa Oumar

KZN Department of Agriculture and Rural Development, South Africa,
zak.oumar@kzndard.gov.za

<http://dx.doi.org/10.4314/sajg.v5i2.9>

Abstract

Lantana camara is a significant weed in South Africa which is causing severe impacts on agriculture by reducing grazing areas. This study assessed the potential of the SPOT 6 multispectral sensor and two broadband vegetation indices (NDVI and SR) for detecting and mapping *Lantana camara* in a community grazing land in KwaZulu-Natal, South Africa. The SPOT 6 bands and vegetation indices successfully classified *Lantana camara* with an overall accuracy of 75% on an independent test dataset using the random forest algorithm. Furthermore, it was tested if the random forest model based on variable importance (VIP) could improve the classification accuracy using the best subset of bands and indices. A backward feature elimination technique was used to select the best subset of VIP bands and indices to improve the classification. By eliminating SPOT bands 1 and 4 which yielded the lowest VIP scores the random forest model improved the classification accuracy to 83.33% on an independent test dataset. The study indicates the potential of satellite remote sensing in weed detection and mapping in South Africa using readily available multispectral data to assist poorer communities in grazing management.

1. Introduction

Lantana camara is a significant weed in South Africa causing severe impacts on agriculture and natural ecosystems (Day *et al.*, 2003). The weed is characterised as a low, erect or subscandent, vigorous shrub which can grow up to 2-4m in height (Sankaran, 2008; Walton, 2006). The weed grows in individual clumps or as dense thickets and suppresses the growth of indigenous plants (Sharma *et al.*, 2005). The plants allelopathic qualities can reduce the vigour of indigenous plant species and reduce the biodiversity of natural ecosystems (Sharma *et al.*, 2005; Walton, 2006). *Lantana* is able to climb up to 15m with the support of other vegetation (Sankaran, 2008) and produces a greenish blue-black fruit which is consumed by birds and animals (Walton, 2006). Germination of the seed is improved when the seed passes through the digestive system of birds and animals and this contributes to the spread of the weed

(Sankaran, 2008). *Lantana* has a significant impact on agriculture by reducing grazing areas, reducing productivity of crops and in dense stands the capacity of the soil to absorb water is reduced therefore potentially increasing runoff and promoting the risk of soil erosion (Cilliers, 1983; Sankaran, 2008). Control of *Lantana camara* is achieved by mechanical clearing and hand pulling which is often utilised for small areas whereas fire is used to control large areas of invasion (Day *et al.*, 2003; PPRI, 1997). Chemical control is also used and is achieved by implementing herbicides which have been registered for the control of *Lantana camara*. Biological control agents such as the *Teleionemia scrupulosa*, a leaf sap sucking bug and *Octotoma scabripennis*, a leaf mining beetle have also been used to control the invasion of *Lantana camara* in South Africa (PPRI, 1997). At present these insects contribute towards the suppression of the growth of *Lantana camara*.

In KwaZulu-Natal (KZN), South Africa, the spread of *Lantana camara* is causing significant damage to community grazing lands. The KZN Department of Agriculture have implemented community landcare projects to remove high infestation rates of *Lantana camara*. These projects are community upliftment programmes which increase job creation and involve the clearing of *Lantana camara* which has reduced the grazing capacity and palatable species in these areas. There is an urgent need to accurately locate the spatial location of *Lantana camara* in order to assist community members with clearing projects and to prevent exorbitant labour costs of field assessments in locating the invasive weed. Furthermore, by quantifying the amount of *Lantana camara* present in these areas will assist with precise budgeting in terms of the quantities of chemicals needed and the necessary equipment required for the removal of the weed.

Traditional methods for identifying invasive weeds involve field surveys and GPS mapping which produce high accuracies but are often financially, technically and logistically impractical for many managers (Cooksey and Sheley, 1998). Several studies have shown the utility of hyperspectral remote sensing technology to accurately detect and map invasive species (Eddy *et al.*, 2014; Jay *et al.*, 2009; Lawrence *et al.*, 2006). Hyperspectral sensors which collect detail spectral information are able to pick up slight changes in vegetation (Oumar and Mutanga, 2011) and are able to detect invasive species with high accuracies. Lawrence *et al.*, (2006) successfully mapped leafy spurge and spotted knapweed using 128-band hyperspectral imagery with overall accuracies ranging from 84% to 86% using the random forest algorithm. Jay *et al.*, (2009) used low cost hyperspectral imagery in conjunction with the random forest algorithm to detect leafy spurge with accuracies ranging from 72% to 95%. More recent results by Eddy *et al.*, (2014) showed the potential of hyperspectral imagery in classifying weeds with accuracies ranging from 88% to 94% using a neural network classification algorithm. Atkinson *et al.*, (2014) showed classification accuracies of 93% using AISA Eagle hyperspectral data

and support vector machines to map bugweed in forestry plantations in South Africa. Peerbhay *et al.*, (2015) also assessed the utility of AISA Eagle hyperspectral data to discriminate bugweed from a variety of forest species in South Africa. While the use of hyperspectral sensors have become more advantageous owing to the wealth of spectral information, hyperspectral data may represent an oversampled dataset, which results in high data dimensionality and redundant wavebands that may be irrelevant for detecting alien invasive plants of interest (Odindi *et al.*, 2014; Peerbhay *et al.*, 2016). Often a major challenge in obtaining hyperspectral data in South Africa is the huge cost of the imagery (Oumar and Mutanga, 2010).

In contrast to hyperspectral imagery, multispectral imagery is more cost effective and readily available in South Africa. Multispectral systems collect data in three to six spectral bands within the visible and mid-infrared regions of the electromagnetic spectrum and have been utilised for invasive alien species mapping (Odindi *et al.*, 2014; Schmidt *et al.*, 2010). The recent signing of the licence agreement in March 2015 between the South African National Space Agency (SANSA) and Airbus Defence and Space (ADS), ensures the availability of SPOT 6 and 7 multispectral datasets in South Africa. Several studies have shown the utility of SPOT data in detecting and mapping invasive species (Odindi *et al.*, 2014; Schmidt *et al.*, 2010). Very recent results by Odindi *et al.*, (2014) showed the utility of WorldView-2 and SPOT-5 images in mapping bracken fern in South Africa with overall classification accuracies of 84.72% and 72.22% using the random forest algorithm. Schmidt *et al.*, (2010) also showed the potential of SPOT 5 multispectral imagery in mapping aquatic weeds in a river system in Australia with good overall accuracies. SPOT 6 and 7 is an improvement of the previous SPOT 5 imagery with spatial resolutions of 1.5m for panchromatic scenes and 6m for multispectral scenes (ADS, 2016). With an improved spatial resolution and strategically placed spectral bands this sensor offers the potential to detect and map *Lantana camara* in the community grazing lands in South Africa. Advanced and robust classifiers such as the random forest algorithm which have shown promising results in invasive species modelling (Lawrence *et al.*, 2006; Odindi *et al.*, 2014) offer potential in *Lantana camara* classification and mapping. It is against this background that the study aimed to assess the utility of SPOT 6 multispectral imagery in conjunction with the random forest algorithm to detect and map *Lantana camara* in a community landcare project in KwaZulu-Natal, South Africa.

2. Study Area

The study area is located in the Ugu district Municipality of KwaZulu-Natal South Africa and covers an area of approximately 3km². The area receives an average annual rainfall of about 950 mm (Schulze and Lynch, 2007) and is situated at an altitude range of about 120 m above sea level (Schulze and Horan, 2007). The mean annual temperature of the area is 20

degrees Celsius (Schulze and Maharaj, 2007). The vegetation types of the area consist of bushed grassland and bush land. The land is used by the indigenous local community for grazing their animals and for beef production. Livestock that is produced is sold to the local communities. However, large areas are covered by *Lantana camara* which is reducing the grazing potential of the area. Figure 1 shows a map of the study area.

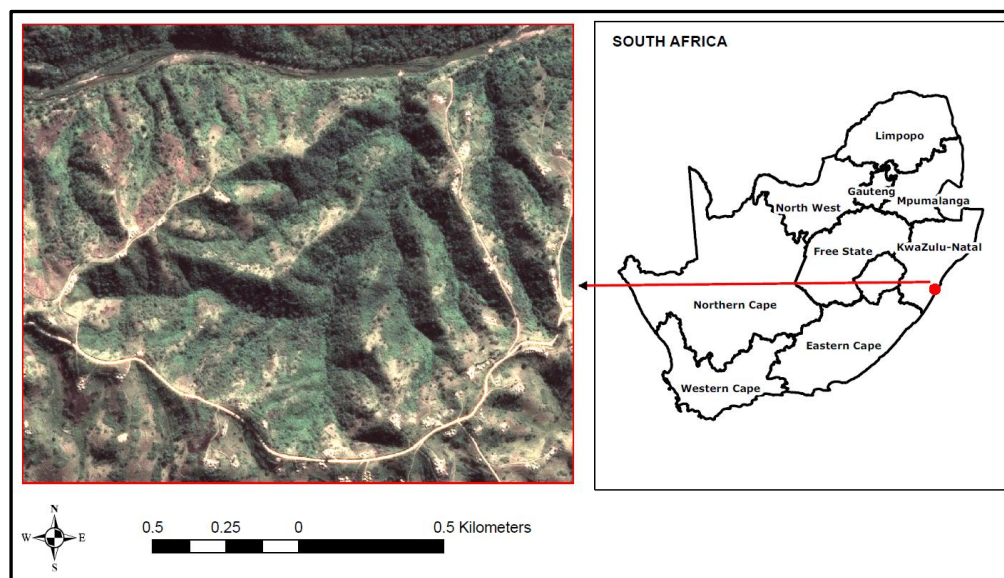


Figure 1. Study Area

3. Methods

3.1. SPOT 6 Imagery

One scene of SPOT 6 imagery covering the study area with an image pass of 12th June 2013 was acquired from the South African National Space Agency. SPOT 6 was launched in September 2012 and contains 5 spectral bands. Table 1 indicates the spectral and spatial resolution of each of these bands.

Table 1. Spectral and spatial resolution of SPOT 6

Spectral Bands	Spectral Range (Micrometres)	Spatial Resolution (Metres)
Panchromatic	0.450-0.745	1.5
Blue	0.450-0.520	6
Green	0.530-0.590	6
Red	0.625-0.695	6
Near-Infrared	0.760-0.890	6

The SPOT 6 scene was orthorectified and projected to the UTM Zone 36S by SANSA. Radiometric correction of the image was carried out by converting the digital numbers to top of atmosphere reflectance using ENVI 5. Furthermore two commonly used broadband

vegetation indices (Equation 1 and 2) were calculated, the Normalized Difference Vegetation Index (NDVI) (Rouse *et al.*, 1973) and the Simple Ratio (SR) (Jordan, 1969) to test its potential in combination with the SPOT 6 bands in detecting and mapping *Lantana camara*. The NDVI uses the red and near-infrared band of the electromagnetic spectrum to assess changes in vegetation phenology as it uses the highest absorption and reflectance of the chlorophyll region. The SR index is the ratio of the near-infrared and red bands and is used to assess changes in green vegetation cover.

$$\text{NDVI} = \frac{\text{NIR} - \text{RED}}{\text{NIR} + \text{RED}} \quad [1]$$

$$\text{SR} = \frac{\text{NIR}}{\text{RED}} \quad [2]$$

3.2. Sampling

Field sampling was carried out by generating forty random 6 x 6m plots over the multispectral image. The 6 x 6m plot size was chosen to ensure that the entire multispectral pixel is covered by *Lantana camara*. Twenty of the 6 x 6m plots were completely covered by *Lantana camara* infestation and the remaining 20 plots had no *Lantana camara* present (Figure 2). The plots were assessed in field for the presence and absence of *Lantana camara*. The forty sampling plots were randomly divided in two sets whereby 70% of the dataset was used for training the model and the remaining 30% was used for validation. Table 2 shows the proportion of the training and test data plots.

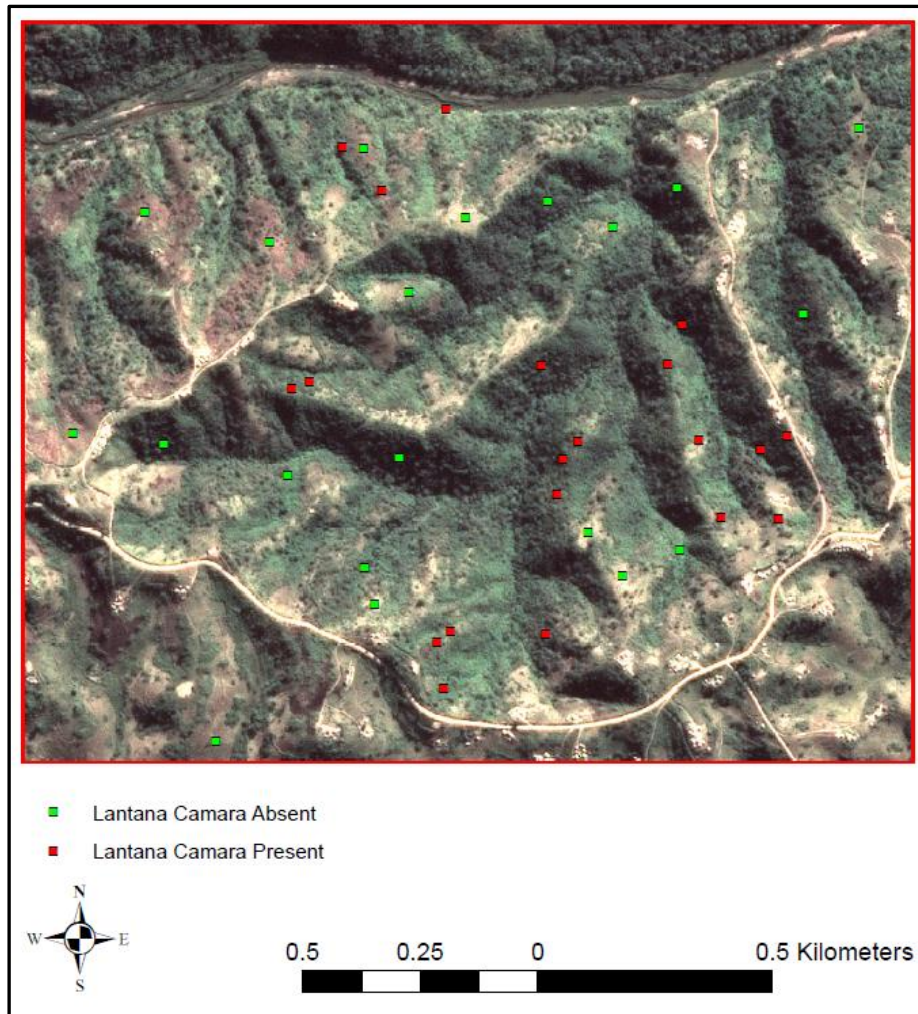


Figure 2. *Lantana camara* sampling plots

Table 2. Proportion of training and test datasets

	<i>Lantana camara</i> present	<i>Lantana camara</i> absent
Training (70%)	14	14
Test (30%)	6	6
Total	20	20

3.3. Random Forest Algorithm

Spectral reflectance values were extracted from the forty sampling plots and were input into the random forest algorithm (Breiman, 2001) in STATISTICA (statistical software package). The random forest algorithm is an ensemble learning technique developed by Breiman (2001) to improve the classification and regression trees method by combining a large set of decision trees. The algorithm uses a bagging (bootstrap aggregation) operation where multiple classification trees are constructed based on a random subset of samples derived from approximately two thirds of the training data. The multiple classification trees then vote by

plurality on the correct classification (Breiman, 2001; Lawrence *et al.*, 2006). The one third samples not used in the bootstrap samples are referred to as the out-of-bag (OOB) sample. The OOB sample can be used to estimate the misclassification error and the variable importance.

The random forest model was optimised using the training dataset for four different classification trees of 25, 50, 75 and 100 on all 6 (mtry) of the SPOT 6 bands and vegetation indices. The lowest OOB error was used to determine the optimal mtry value which was then used in the subsequent analysis.

3.4. Variable Importance

The random forest algorithm calculates the importance of each explanatory variable in the model and determines which variable or set of variables are most relevant based on the classification error. In order to determine the optimal number of variables which resulted in the lowest classification error a backward feature elimination method was applied on the variable importance (VIP) variables (Ismail *et al.*, 2010). The backward feature elimination method is a common method applied in remote sensing studies to filter important variables and involves building random forest models and then progressively eliminates the VIP variable with the smallest importance (Ismail *et al.*, 2010; Odindi *et al.*, 2014). The smallest subset of variables with the lowest OOB error was then used to classify *Lantana camara*. The random forest model was implemented in EnMap Box v. 1.4, to create a classified map of *Lantana camara* over the study area.

3.5. Accuracy

The overall accuracy of the classified map was validated using an error matrix constructed from the independent 30% dataset. Furthermore, the KHAT statistic (k) which is a measure of accuracy was used to determine if the overall classification was better than if it was classified by a random classifier. KHAT values range from -1 to +1, and if the values are 1 or close to 1 then there is perfect agreement between the test and training datasets (Congalton and Green, 1999; Lillesand *et al.*, 2004).

4. Results

4.1. Mean *Lantana camara* Reflectance

The mean reflectance of *Lantana camara* is shown in Figure 3. The reflectance indicated a normal spectral vegetation curve with low reflectance in the visible spectrum and an increase in the near-infrared reflectance. Low reflectance in the visible spectrum is due to the plant absorbing energy from the sun as compared to the near-infrared which increases in reflectance due to low absorption.

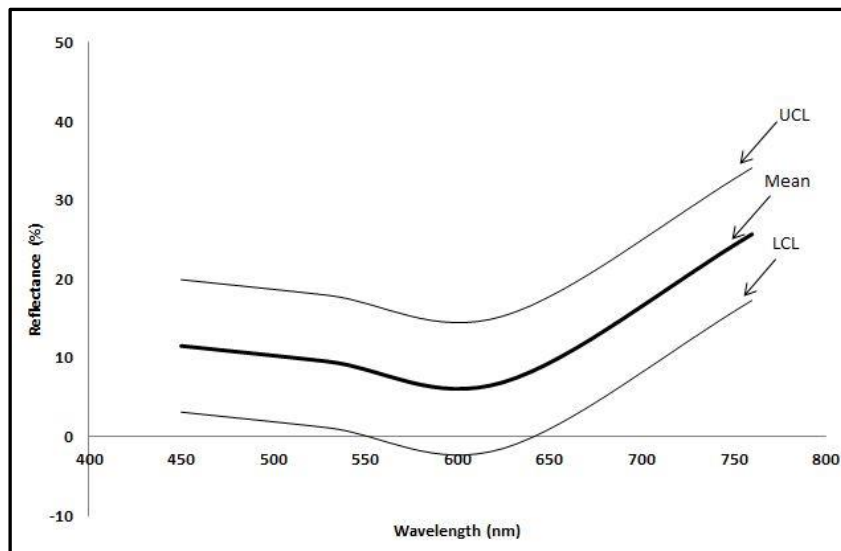


Figure 3. *Lantana camara* reflectance from SPOT 6 bands (n=20) showing the mean, 95% upper confidence limit (UCL), and 95% lower confidence limit (LCL) of the reflectance.

4.2. Random Forest Optimisation

The random forest optimisation (Table 3) indicated that the OOB error varied from 17.86% to 28.57%. The lowest OOB error was obtained using 100 trees with 4 mtrys. An mtry value of 4 and an ntree value of 100 were then used for the classification.

Table 3. Random forest optimisation and classification

Ntree value	25	50	75	100
Minimum OOB error	28.57%	25.00%	21.43%	17.86%
Optimal mtry value	3	1	3	4
OOB error (mtry=4)	32.14%	35.71%	35.71%	17.86%

4.3. Random Forest Classification

The confusion matrix in Table 4 indicates the performance of the random forest algorithm in classifying the presence and absence of *Lantana camara* on the independent 30% test dataset. The random forest model classified the presence and absence of *Lantana camara* with an overall accuracy of 75% with a k statistic of 0.50. The producer accuracy was relatively high at 80% whereas the user accuracy was at 66.67%.

Table 4. Confusion matrix

Class	<i>Lantana camara</i> present	<i>Lantana camara</i> absent	Row total
<i>Lantana camara</i> present	4	2	6
<i>Lantana camara</i> absent	1	5	6
Column total	5	7	12
Producer accuracy	= 80%	Overall accuracy	= 75%
User accuracy	= 66.67%	Kappa	= 0.50

4.4. Random Forest and Variable Importance

The variable importance in Figure 4 shows the importance of all the mtry values in the classification. Important variables have a high score indicating that the performance of the classification will deteriorate if that variable is no longer available in the model. The VIP bar graph indicated that the NDVI and SR were the most significant variables in the random forest model with scores of 1 and 0.97 followed by SPOT bands 2 (Green) and 3 (Red) with scores of 0.70 and 0.63 respectively.

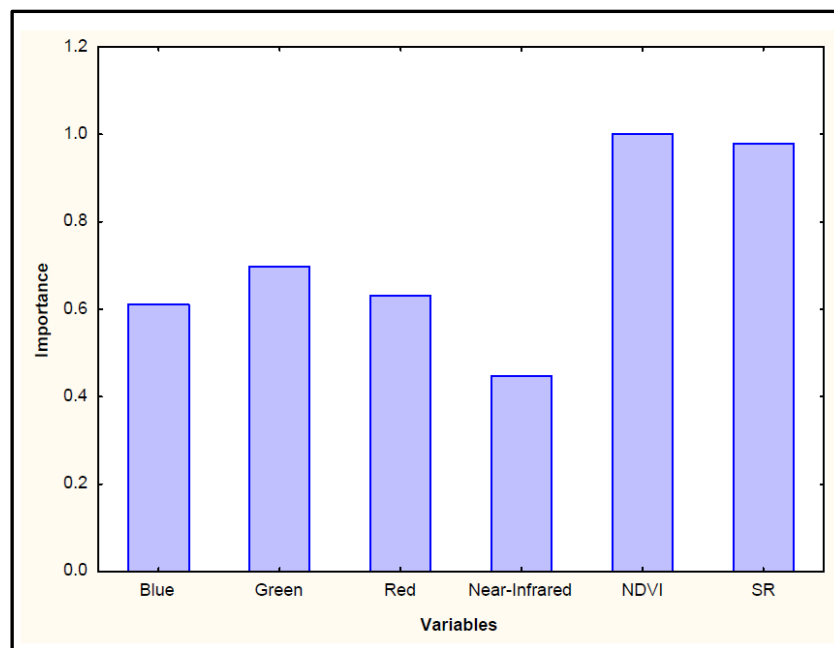


Figure 4. Random forest VIP

The backward feature elimination method was used to select the best subset of VIP variables to improve the classification. By eliminating the SPOT bands 1 (Blue) and 4 (Near-Infrared) which yielded the lowest VIP scores, the random forest model produced a classification accuracy of 83.33% compared to 75% using all the bands. The confusion matrix in Table 5 indicates the performance of the optimal VIP variables in classifying *Lantana camara* on the independent test dataset. The k statistic also improved from 0.50 to 0.66 and the producer and user accuracy were relatively high at 75% and 100%. Figure 5 shows the classified map over

the study area using the random forest optimal VIP variables. The rivers and roads were masked out of the classification.

Table 5. Confusion matrix

Class	<i>Lantana camara</i> present	<i>Lantana camara</i> absent	Row total
<i>Lantana camara</i> present	6	0	6
<i>Lantana camara</i> absent	2	4	6
Column total	8	4	12
Producer accuracy	= 75%	Overall accuracy	= 83.33%
User accuracy	= 100%	Kappa	= 0.66

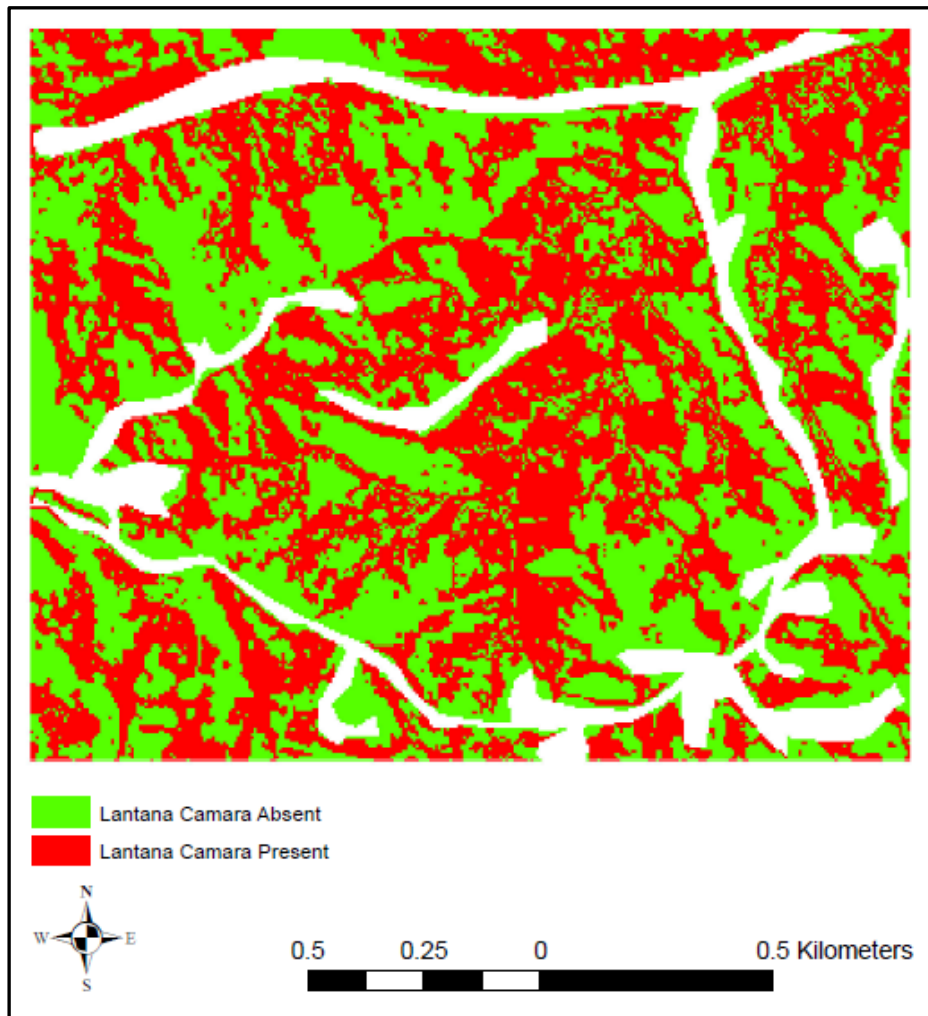


Figure 5. Random forest classified map

5. Discussion

5.1 SPOT 6 Imagery and *Lantana Camara* Detection and Mapping

The ability to accurately locate and map invasive species provides valuable information for targeting eradication, identifying management actions and monitoring the effectiveness of mitigation efforts (Schmidt *et al.*, 2010). Maps of *Lantana camara* outbreak provide essential information to the local communities in order to prioritise eradication strategies, provide rehabilitation methods to the landscape and to identify land available for grazing their animals. The high spatial resolution of the SPOT 6 satellite together with the shorter revisit time of 3 to 4 days makes this sensor well suited for large scale weed detection as it offers an alternative to traditional mapping approaches based on fieldwork and aerial photography. The study indicated the potential of SPOT 6 imagery and broadband vegetation indices for detecting and mapping *Lantana camara* with an overall accuracy of 75% on an independent test dataset. By utilising the best subset of VIP predictors the classification accuracy improved to 83.33 %. Similar results were obtained by Odindi *et al.*, (2014) who mapped bracken fern with WorldView-2 data and received an overall accuracy of 91.67% using selected vegetation indices based on VIP as compared to 82.33% using all vegetation indices.

The NDVI and SR indices calculated from the SPOT 6 satellite data were identified as the most important variables for the detection of *Lantana camara*. Vegetation indices calculated from the red and near-infrared region respond to changes in plant phenology (Penuelas and Filella, 1998), and thus could explain its importance in detecting *Lantana camara* from surrounding vegetation. The study further indicates the robustness of the random forest algorithm in identifying key variables which enhance classification results. The applicability of the new SPOT 6 sensor in successfully detecting and mapping *Lantana camara* with high accuracies makes it a significant data source in South Africa for weed mapping and monitoring. This study shows the integral part satellite remote sensing plays in assisting poorer communities in veld and grazing management in South Africa.

6. Conclusion

The aim of this study was to assess the potential of the SPOT 6 multispectral data in conjunction with the random forest algorithm for detecting and mapping *Lantana camara* in a community grazing project. The results show the capability of the SPOT 6 sensor in successfully classifying *Lantana camara* with an overall accuracy of 75% on an independent test dataset. The classification accuracy was further strengthened to 83.33% using only the best subset of variables and by applying the random forest VIP model with a backward elimination technique. The research corroborates findings by other studies who have applied similar techniques (Odindi *et al.*, 2014) for the detection of invasive alien plants in South Africa using multispectral imagery. The result is significant for invasive species mapping and monitoring in South Africa using readily available SPOT 6 multispectral datasets.

7. References

ADS 2016, SPOT 6/SPOT 7 Technical Sheet, Airbus Defence and Space, France, viewed December 2015, < <http://www.geo-airbusds.com/en/147-spot-6-7-satellite-imagery> >.

Atkinson, JT, Ismail, R, & Robertson, M 2014, 'Mapping bugweed (*Solanum mauritianum*) infestations in *pinus patula* plantations using hyperspectral imagery and support vector machines', *IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing*, vol. 7, pp. 17-28.

Breiman, L 2001, 'Random forests', *Machine Learning*, vol. 45, pp. 5-32.

Cilliers, CJ 1983, 'The weed, *Lantana camara*, and the insect natural enemies imported for its biological control into South Africa', *Journal of the Entomological Society of Southern Africa*, vol. 46, pp. 131-138.

Congalton, RG & Green, K, 1999, *Assessing the accuracy of remotely sensed data: Principles and practices*, Lewis, Boca Raton, FL.

Cooksey, D & Sheley, RL 1997, 'Noxious weed survey and mapping system', *Rangelands*, vol. 19, pp. 20-23.

Day, MD, Wiley, CJ, Playford, J & Zalucki, MP, 2003, *Lantana: Current Management Status and Future Prospects*, Australian Centre for International Agricultural Research, Canberra, Australia.

Eddy, PR, Smith, AM, Hill, BD, Peddle, DR, Coburn, CA & Blackshaw, RE 2013, 'Weed and crop discrimination using hyperspectral image data and reduced bandsets', *Canadian Journal of Remote Sensing*, vol. 39, pp. 481-490.

Ismail, R, Mutanga, O & Kumar, L 2010, 'Modeling the Potential Distribution of Pine Forests Susceptible to *Sirex noctilio* Infestations in Mpumalanga, South Africa', *Transactions in GIS*, vol. 14, pp. 709-726.

Jay, S, Lawrence, R, Repasky, K & Keith, C 2009, 'Invasive species mapping using low cost hyperspectral imagery', in *ASPRS 2009 Annual Conference Proceedings*, Baltimore, Maryland, 2009.

Jordan, CF 1969, 'Derivation of leaf area index from quality of light on the forest floor', *Ecology*, vol. 50, pp. 663-666.

Lawrence, RL, Wood, SD & Sheley, RL 2006, 'Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (RandomForest)', *Remote Sensing of Environment*, vol. 100, pp. 356-362.

Lillesand, T, Kiefer, R & Chipman, J, 2004, *Remote sensing and image interpretation*, John Wiley and Sons, New York.

Odindi, J, Adam, E, Ngubane, Z, Mutanga, O & Slotow, R 2014, 'Comparison between WorldView-2 and SPOT-5 images in mapping the bracken fern using the random forest algorithm', *Journal of Applied Remote Sensing*, vol. 8, pp. 083527.

Oumar, Z & Mutanga, O 2010, 'Predicting plant water content in *Eucalyptus grandis* forest stands in KwaZulu-Natal, South Africa using field spectra resampled to the Sumbandila Satellite Sensor', *International Journal of Applied Earth Observation and Geoinformation*, vol. 12, pp. 158-164.

Oumar, Z & Mutanga, O 2011, 'The potential of remote sensing technology for the detection and mapping of *Thaumastocoris peregrinus* in plantation forests', *Southern Forests: A Journal of Forest Science*, vol. 73, pp. 23-31.

Peerbhay, KY, Mutanga, O & Ismail, R 2015, 'Random forests unsupervised classification: The detection and mapping of *solanum mauritianum* infestations in plantation forestry using hyperspectral data', *IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing*, vol. 8, pp. 3107 – 3122.

Peerbhay, KY, Mutanga, O & Ismail, R 2015, 'The identification and remote detection of alien invasive plants in commercial forests: An Overview', *South African Journal of Geomatics*, vol. 5, pp. 49-67.

Penuelas, J, Filella, L, 1998, 'Visible and near-infrared reflectance techniques for diagnosing plant physiological status', *Trends in Plant Science*, vol. 3, pp. 151–156.

PPRI, 1997, *Control of Lantana camara*, Plant Protection Research Institute, Agricultural Research Council, Pietermaritzburg, South Africa.

Rouse, JW, Haas, RH, Schell, JA & Deering, DW 1973 'Monitoring Vegetation Systems in the Great Plains with ERTS', *Third ERTS Symposium*, Washington, DC, 1973.

Sankaran, KV, 2008, *Invasive Pest Fact Sheet Lantana camara*, Asia - Pacific Forest Invasive Species Network, Kerala, India.

Schmidt, M & Witte, C 2010, 'Monitoring aquatic weeds in a river system using SPOT 5 satellite imagery', *Journal of Applied Remote Sensing*, vol. 4, pp. 043528.

Schulze, RE & Horan, MJC 2007, 'Altitude and Relative Relief', In: Schulze, RE (ed.), *South African Atlas of Climatology and Agrohydrology*, Water Research Commission, Pretoria, RSA.

Schulze, RE & Lynch, SD 2007, 'Annual Precipitation', In: Schulze, RE (ed.), *South African Atlas of Climatology and Agrohydrology*, Water Research Commission, Pretoria, RSA.

Schulze, RE & Maharaj, M 2007, 'Mean Annual Temperature', In: Schulze, RE (ed.), *South African Atlas of Climatology and Agrohydrology*, Water Research Commission, Pretoria, RSA.

Sharma, GP, Raghubanshi, AS & Singh, JA 2005, 'Lantana invasion: An overview', *Weed Biology and Management*, vol. 5, pp. 157-165.

Walton, C, 2006, *Lantana camara (shrub)*, Global Invasive Species Database, Auckland, viewed December 2015, <<http://www.issg.org/database/species/ecology.asp?si=56&fr=1&sts=&lang=EN>>.