⊞ AOSIS

Mapping alien and indigenous vegetation in the KwaZulu-Natal Sandstone Sourveld using remotely sensed data

CrossMark
←click for updates

Authors:
John Odindi[1]
Onisimo Mutanga[1] ⓘD
Mathieu Rouget[1] ⓘD
Nomcebo Hlanguza[1] ⓘD

Affiliations:
[1]School of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal, South Africa

Corresponding author:
John Odindi,
odindi@ukzn.ac.za

Background: The indigenous KwaZulu-Natal Sandstone Sourveld (KZN SS) grassland is highly endemic and species-rich, yet critically endangered and poorly conserved. Ecological threats to this grassland ecosystem are exacerbated by encroachment of woody plants, with severe negative environmental and economic consequences. Hence, there is an increasing need to reliably determine the extent of encroached or invaded areas to design optimal mitigation measures. Because of inherent limitations that characterise traditional approaches like field surveys and aerial photography, adoption of remotely sensed data offer reliable and timely mapping of landscape processes.

Objectives: We sought to map the distribution of woody vegetation within the KZN SS using remote sensing approaches.

Method: New generation RapidEye imagery, characterised by strategically positioned bands, and the advanced machine learning algorithm Random Forest (RF) were used to determine the distribution and composition of alien and indigenous woody vegetation within the KZN SS.

Results: Results show that alien and indigenous encroachment and invasion could be mapped with over 86% accuracy whilst the dominant indigenous and alien tree species could be mapped with over 74% accuracy. These results highlight the potential of new generation RapidEye satellite data in combination with advanced machine learning technique in predicting the distribution of alien and indigenous woody cover within a grassland ecosystem. The successful discrimination of the two classes and the species within the classes can be attributed to the additional strategically positioned bands, particularly the red-edge in the new generation RapidEye image.

Conclusion: Results underscore the potential of new generation RapidEye satellite data with strategically positioned bands and an advanced machine learning algorithm in predicting the distribution of woody cover in a grassland ecosystem.

Introduction

The grassland biome is one of the most dominant land cover types on earth, covering approximately 52.5 million km^2 (40.5%) of the global terrestrial landscape – without Greenland and Antarctica (Carlier *et al.* 2009; World Resources Institute 2000). In South Africa, the grassland is the second largest and most diverse biome, covering 16.5% of the country (Matsika 2007). In the KwaZulu-Natal province, a significant portion of the landscape is characterised by the indigenous KwaZulu-Natal Sandstone Sourveld (KZN SS), a veld type dominated by a diversity of short grass species, shrubs, legumes and trees (Rutherford *et al.* 2006). However, the veld has been steadily diminishing. Jewitt (2011), for instance, notes that only 11.4% of the KZN SS was in its natural habitat as at 2008. The eThekwini Municipality (Figure 1), which has the port city of Durban, is the most densely populated area within the province with a population of about 3.5 million on about 2300 km^2 (eThekwini Municipality 2012). Increasing immigration, agriculture, land tenure challenges and radial growth from the city have increasingly put pressure on the city's natural landscape (Roberts *et al.* 2012). Hence about 73% of the veld has been lost to agriculture and physical development. A paltry 116 ha (0.74%) of the veld is formally protected (eThekwini Municipality 2012).

The remnant KZN SS covers a portion of the municipality and anchors soil, contributes to carbon sequestration, sustains the agricultural sector, is a habitat for a variety of endangered species and is a source of traditional medicine (Martindale 2007; Rutherford *et al.* 2006). It is diverse in graminoids and herbaceous species and hosts an array of endemic plants that play an integral role in the
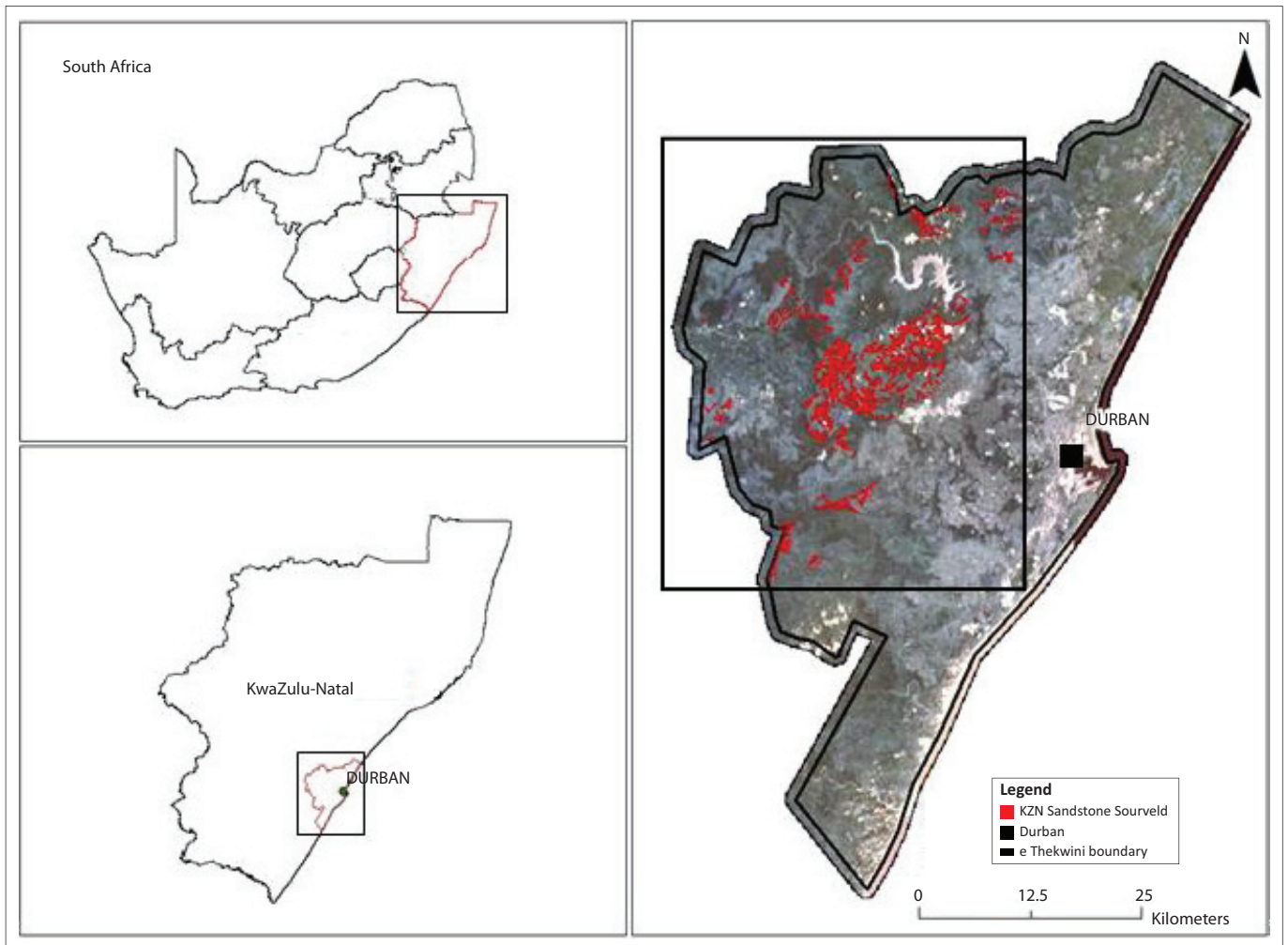
Source: Authors' own work

FIGURE 1: The study area.

grassland's ecosystem (Rutherford *et al.* 2006). However, recent studies have shown that the KZN SS is threatened by encroachment of woody plants. Such encroachment is known to significantly compromise a grassland's quality and quantity; for instance, an increase in shrub or tree density may reduce grass biomass, density and cover (Briggs, Schaafsma & Trekov 2007; Van Auken 2009). Generally, species richness and composition is negatively altered as woody plants begin to dominate the landscape (Van Auken 2009). The problem is further exacerbated if the species are invasive (Lalla 2014). Hence, timely and cost-effective mapping of the indigenous and invasive encroachment is necessary for designing appropriate encroachment mitigation measures within the municipality's Municipal Adaptation Plan (MAP). Specifically, understanding the proportion and spatial configuration of alien versus indigenous woody encroachment is critical in discerning the severity and type of woody vegetation encroachment. However, within the municipality, spatial distribution of indigenous and alien woody plants within the KZN SS remain largely unexplored.

To date, studies on woody encroachment have commonly been explored using field surveys, field-based knowledge and interpretation and analysis of hard-copy maps and aerial photographs, amongst others (Shekede, Murwira & Masocha 2015; Yuan *et al.* 2005). However, although relatively accurate, these approaches require intensive field work and ancillary data analysis, which is labour intensive and susceptible to human error (Shekede *et al.* 2015). Furthermore, these approaches are often time consuming, and therefore impractical for large-scale implementation, and commonly lack the required geometric accuracy (Mansour, Mutanga & Everson 2012; Xie, Sha & Yu 2008). Remotely sensed data sets and approaches, in comparison to the abovementioned approaches, offer a more practical and economical means of classifying and quantifying vegetation characteristics and density (Mansour *et al.* 2012). The recent wide adoption of remote sensing approaches can be attributed to advances in sensor technology, proliferation of image data sets and advances in software and hardware capabilities (Rogan & Chen 2004). For instance, improvements in satellite-based remote sensing technologies have led to the acquisition of imagery characterised by finer spatial and higher spectral resolutions, necessary for improving classification accuracies (Adelabu, Mutang & Adam 2014).

A number of multispectral data (e.g. Landsat) have been adopted in land cover mapping on the basis of their long

archival data (Odindi, Mhangara & Kakembo 2012). However, their lower spectral resolution (less than 11 bands) limits their capabilities to discriminate surface types characterised by subtle reflectance variation (Melgani & Bruzzone 2004). Munyati, Shaker and Phasha (2011) further note that adoption of readily available remotely sensed data, such as the Advanced Very High Resolution Radiometer (NOAAH-AVHRR) and Moderate Resolution Imaging Spectroradiometer (MODIS) in heterogeneous landscapes are limited by their coarse spatial resolution and the mixed pixel phenomenon. In contrast, hyperspectral sensors, with higher spatial and spectral resolution, have been identified as an alternative for reliable land-use-land-cover mapping. In contrast to commonly used multispectral imagery, hyperspectral imagery are characterised by narrow band widths in the visible, near-infrared (NIR) and shortwave-infrared sections of the electromagnetic spectrum that facilitate better discrimination of surface types (Abdel-Rahman *et al*. 2014a; Mansour *et al*. 2012; Peerbhay, Mutanga & Ismail 2013). However, hyperspectral image data are relatively expensive, require time consuming pre-processing, suffer from multi-collinearity and require large storage space, amongst other things (Adelabu *et al*. 2014; Eitel *et al*. 2007, Kiala *et al*. 2016). Recently, improvements in remote sensing technologies have led to acquisition of imagery characterised by finer spatial and higher spectral resolutions (Adelabu *et al*. 2014). New generation sensors such as RapidEye (5 m, 5 bands), Worldview-2 (0.3–3.7 m, 16 bands), Sentinel (8–10 m, 7–12 bands) and Quickbird (2.4 m, 4 bands), for instance, are characterised by higher spatial and strategically positioned bands. These characteristics can compensate for the shortfalls experienced by existing hyperspectral and traditional medium or low resolution sensors like Landsat, SPOT and ASTER (Munyati *et al*. 2011; Ojoyi *et al*. 2016). Furthermore, sensors like RapidEye are characterised by additional red-edge band, which facilitates discrimination of subtle variation in vegetation (Schuster, Forster & Kleinschmit 2012; Tigges, Lakes & Hostert 2013). It is therefore necessary to test the use of new generation sensors, such as the RapidEye, in determining the presence of woody vegetation within the KZN SS.

Whereas higher spectral resolution is valuable for discriminating subtle reflectance variability on a landscape, it does not guarantee higher classification accuracy as this is further dependent on the classifier adopted (Lu & Weng 2007; Mountrakis, Im & Ogole 2011). A number of classifiers that include classical pattern recognition approaches like maximum likelihood, minimum distance to mean and nearest neighbour as well as discriminant analysis have been adopted for land cover mapping. However, Lu and Weng (2004) note that these classifiers are often unable to effectively deal with complex landscapes and the mixed pixel phenomenon, compromising mapping accuracy. To deal with these limitations, more advanced and robust machine learning classification algorithms have recently emerged (Sesnie *et al*. 2010; Watanachaturaporn, Arora & Varshney 2008). Specifically, the Random Forest (RF) algorithm has shown great potential in vegetation mapping (Adelabu *et al*. 2013; Lawrence, Wood & Sheley 2006; Naidoo *et al*. 2012). First developed by Breiman (2001), the RF is a non-parametric statistical algorithm that can handle discrete and continuous data sets and has been adopted in a wide range of remotely sensed data sets that include multispectral (Pal 2005), hyperspectral (Ham *et al*. 2005), LIDAR (Guo *et al*. 2011), synthetic aperture radar (Loosevelt *et al*. 2012) and imagery from aerial platforms (Chapman *et al*. 2009). A number of studies have adopted the algorithm to predict the occurrence of trees in grasslands and savannah landscapes. Naidoo *et al*. (2012), for instance, used the RF algorithm to classify tree species within the greater Kruger National Park in South Africa whilst Lawrence *et al*. (2006) assessed the capabilities of the RF algorithm in identifying alien invasive species in Montana, United States of America. Whereas the abovementioned studies have attempted to discriminate woody vegetation in grasslands using remotely sensed data, none, to our knowledge, has sought to distinguish the presence of alien from indigenous species within grasslands in a heterogeneous urban landscape. In this study, we aim to map alien and indigenous woody cover distribution within the KZN SS using the RF algorithm on RapidEye imagery.

The study area

The general area is located within a subtropical climate, characterised by hot and humid summers and sunny and mild winters. Most of the 762 mm per annum of rain is experienced during summer months (October–March) whilst winter months (April–August) are generally dry (Mucina & Rutherford 2006). Average midday temperatures of 27 °C and 21.6 °C are experienced in the summer and winter months, respectively. There is a relatively small difference between summer and winter temperatures. Temperate and moist conditions because of underlying clastic sedimentary sandstone that allows for the percolation of water in the study area are ideal for grasses to flourish (Martindale 2007).

The KZN SS is a unique grassland ecosystem in South Africa, endemic to the KwaZulu-Natal province (Jewitt 2011; Mucina & Rutherford 2006). It is characterised by a matrix of short and diverse grasses, isolated shrubs and woody plants. The grassland often dominates plateau surfaces formed on Natal Group Sandstone resistant to erosion. Soils within the grassland, underlain by the Natal Group Sandstone, are often shallow, infertile and poorly drained (Mucina & Rutherford 2006). Patches of the KZN SS grassland are predominantly found in the south-eastern part of the KwaZulu-Natal Province (Mucina & Rutherford 2006). Within the eThekwini Municipality, patches of the grassland occupy some of the remnant natural landscapes in the western part of the area (Figure 1). The ecosystem is mainly threatened by commercial agriculture (mainly sugarcane and forest plantations), subsistence farming and urbanisation (Mucina & Rutherford 2006). A detailed description of the grassland can be found in Mucina and Rutherford (2006).

Methods

Image acquisition

To cover the entire study area, 12 scenes of RapidEye imagery, detailed in Table 1, were acquired in August 2013 and stitched using ENVI 4.3 software. Geometric and atmospheric corrections and geo-referencing was carried out by the data suppliers. As the focus of this study was on natural landscape, natural areas were extracted from the national land cover map generated from SPOT 5 image with 2.5 m panchromatic and 10 m multispectral bands. The maps were validated using field surveys, aerial photographs and expert advice (Department of Rural Development and Land Reform 2009).

Field data collection

Field data within the KZN SS were collected in September 2014. Stratified purposive non-random sampling was used to select stands dominated by grass, indigenous plants and invasive alien species. This approach was adopted to ensure a representative sample. Using visual categorisation and estimation recommended by Fraser, Abuelgasim and Latifovic (2005), a 10 × 10 m quadrat was established and grass, indigenous and invasive cover estimated and recorded. The 10 × 10 m quadrats were used to ensure that the 5 × 5 RapidEye pixels were accommodated within sampling sites. Stands with over 75% indigenous canopy cover were classified as indigenous whilst those with over 75% alien cover were categorised as alien. A total of 65 alien, 74 indigenous trees and 20 pure grass stands were sampled (Table 2), totalling 159 samples. Using the protocol explained above, a second data set, totalling 153 samples, was also collected in the field where the canopy cover of the five dominant woody species was recorded. The indigenous species considered in this study were *Syzygium cordatum*, *Millettia grandis* and *Strelitzia nicolai* whilst the alien species were *Lantana camara* and *Eucalyptus grandis*. A random 70/30 split (Tables 2 and 3) was applied to the two data sets to generate training and test data. In comparison to other combinations, a number of studies (e.g. Adelabu, Mutang & Adam 2015; Henderson *et al*. 2005; Rogan *et al*. 2008) noted that the 70/30 data split yields the highest classification accuracy and the lowest standard deviation. Co-ordinates were then taken at sampled plots and reflectance extracted from the five RapidEye bands using zonal statistic, a spatial tool within the Spatial Analyst toolbox in ArcGIS 10.3.1 (Environmental Systems Research Institute 2016).

Statistical analysis

Random Forest

The RF ensemble was used for data analysis and optimum parameters input in the EnMAP-Box to classify RapidEye imagery. The ensemble is a machine learning approach established by Breiman (2001) that facilitates better classification and regression of trees (CART) through a combination of a large set of de-correlated decision trees (Lin *et al*. 2010). It benefits from the powerful bagging and

TABLE 1: RapidEye spectral bands and wavelengths.

Band	Name	Wavelength (nm)
Band 1	Blue	440–510
Band 2	Green	520–590
Band 3	Red	630–685
Band 4	Red-edge	690–730
Band 5	Near-infrared	760–850

Source: Authors' own work

TABLE 2: Training and test data of indigenous and alien plots collected within the KZN SS.

Samples	Training (70)	Test (30)	Total
Indigenous	51	23	74
Alien	45	20	65
Grass	14	6	20
Total	**110**	**49**	**159**

Source: Authors' own work

TABLE 3: Training and test data of dominant indigenous and alien tree species collected within the KZN SS.

Samples	Training (70)	Test (30)	Total
Grass	14	6	20
Water berry (*Syzygium cordatum*)	17	8	25
Tickberry (*Lantana camara*)	16	7	23
Umzimbeet (*Millettia grandis*)	14	6	21
Rose gum (*Eucalyptus grandis*)	20	8	28
Natal Wild Banana (*Strelitzia nicolai*)	25	11	36
Total	**106**	**46**	**153**

Source: Authors' own work

random selection process (Lin *et al*. 2010). In the classification process, the RF first builds several binary classification trees (referred to as *ntree*) based on a number of bootstrap samples with replacement extracted from the original observations. An input vector is assigned a tree in the forest. Every tree then provides a classification, regarded as a 'voting' for the class. The forest then determines the classification with the highest number of votes based on tree votes. Samples not accommodated in the bootstrap sample are referred to as out-of-bag (OOB) samples. These samples, typically about 37% of the entire data set, are often used to estimate the misclassification error and variable importance. Thereafter, at each subsequent node, a specific quantity of input variables (referred to as *mtry*), in this case canopy classes, are indiscriminately picked from a random subset of the features and an optimal split determined using the subset of the used features. To guarantee diversity amongst trees and therefore reduce bias, there is no pruning and all trees within the forest are allowed to grow maximally (Breiman 2001; Lin *et al*. 2010).

To improve the accuracy of the classification, the *mtry* and *ntree* require optimisation (Adam *et al* 2014; Breiman 2001). The default *ntree* (number of trees) is set at 500, whilst the *mtry* (number of variables) is the square root of the total number of spectral bands used in the *P*, where *P* equals the number of predictor variables within a data set (Abdel-Rahman *et al*. 2014b; Adelabu *et al*. 2013; Breiman 2001; Naidoo *et al*. 2012). A major advantage of RF is that it is not subject to over-fitting as it can handle irrelevant attributes and ignore noise (Ismail & Mutanga 2010; Naidoo *et al*. 2012;

Peters *et al.* 2007). Furthermore, it is relatively easy to interpret and implement. However, because the algorithm uses thresholds to separate classes, only vertical and horizontal boundaries can be used because one attribute is split at a time (Abdel-Rahman *et al.* 2014b; Breiman 2001). An extended description of the RF can be found in Breiman (2001).

Random Forest optimisation

Optimisation of the RF algorithm was achieved using the two parameters (*ntree* and *mtry*). This process was aimed at determining the optimal parameters required to achieve the most precise classification result (Abdel-Rahman *et al.* 2014a; Adelabu *et al.* 2013; Breiman 2001). The grid search optimisation procedure of the 10 cross-fold validation was implemented to achieve this objective. This approach is the most effective when dealing with few parameters. Furthermore, it is simple to perform, quickly executed and reliable as it considers parameters to be independent. The RF algorithm was executed on the training data to map the presence of alien and indigenous trees and the dominant tree species within the KZN SS grassland.

Map generation

As mentioned, a map of the presence and absence of indigenous woody plant cover was generated in the EnMAP-Box using RF output from RapidEye imagery. The spectral reflectance extracted using field data was used to separate and classify classes at each node. A second and more detailed map showing the dominant indigenous and alien trees was then produced by executing RF using the training data (*n* = 106) and occurrence of alien and other dominant species mapped individually. This was necessary to illustrate the presence of indigenous and alien woody plants.

Accuracy assessment

Accuracy assessment is paramount in any classification process. In this study, a 70% subset (*n* = 110 for the first data set and *n* = 106 for the second data set) of the observed data was used for training and modelling. The remaining 30% (*n* = 49 for the first data set and *n* = 46 for the second data set) was used for model validation. A confusion matrix was used to determine the absence and presence of alien vegetation and the producer's, user's and overall accuracies. Producer's accuracy indicates the probability of a reference pixel being classified correctly. It is determined by dividing a category's number of correct pixels by the category's reference data (Congalton 1991). The user's accuracy indicates the probability that a pixel classified on the map represents that category on the ground. It is determined by dividing the correct number of pixels in a category by the pixels that were actually classified in the category (Congalton 1991). Overall accuracy is the summation of the number of pixels correctly classified divided by the total number of pixels (Congalton 1991).

A quantity and allocation disagreement suggested by Pontius and Millones (2011) was used to determine the classification accuracy. The quantity and allocation of disagreement is defined by Pontius and Millones (2011) as the number of variant pixels in the reference map output in comparison to related maps of imperfect proportions of categories. An extended description, including formulae, can be found in Pontius and Millones (2011).

Results

### Random Forest optimisation and classification

The *mtry* and *ntree* values which produced the least error were selected to classify the presence of woody alien and indigenous plants in the KZN SS after the grid search optimisation procedure. The default number of trees (*ntree* = 500) did not yield an appropriate error. Therefore, the *ntree* was increased to 5000. A value of 2 for the *mtry* produced the smallest OOB error of 18%. Generally, based on the error matrix for the training and test data, the RF algorithm yielded relatively good accuracy (86%) when mapping the existence of indigenous and alien woody cover (Figure 2). These general classes were further classified into dominant indigenous and alien species within the KZN SS. A relatively good accuracy (74%) was achieved when mapping the indigenous species and alien woody cover.

Accuracy assessment

Overall, individual class accuracies for the vegetation types were generally high: 95%, 88% and 82% for grass, indigenous trees and alien vegetation, respectively. The producer's accuracy (PA) for grass, indigenous trees and alien vegetation were 86.54%, 86.07% and 81%, respectively, whilst the user's accuracy (UA) for grass, indigenous trees and alien vegetation were 92.86%, 88% and 82%, respectively. Dominant species *Lantana camara*, *Syzygium cordatum*, *Millettia grandis*, *Eucalyptus grandis* and *Strelitzia nicolai* yielded satisfactory overall accuracies (68%, 75%, 70%, 78% and 80%, respectively). The PA and UA accuracies are shown in Figure 3. Grass had the highest PA (81%) and UA (91%) whilst *Lantana camara* had the lowest PA (62%) and UA (68%) (Figure 3). Low allocation disagreement (17%) and quantity disagreement using the algorithm were obtained.

Discussion

To date, a number of studies (e.g. Belluco, Camuffo & Ferrari 2006; Govender *et al.* 2008; Thenkabail *et al.* 2004) have demonstrated the superiority of hyperspectral sensors in discriminating tree species. This is attributed to their higher spectral resolution, which allows for discrimination of subtle variation in vegetation canopy. However, constraints such as cost, availability and huge data dimensionality have limited their wide adoption. Results in this study have shown that the RapidEye imagery, characterised by fewer but strategically positioned bands like the blue and the red-edge bands, provides a viable alternative to hyperspectral imagery in determining the distribution of alien and indigenous woody cover.
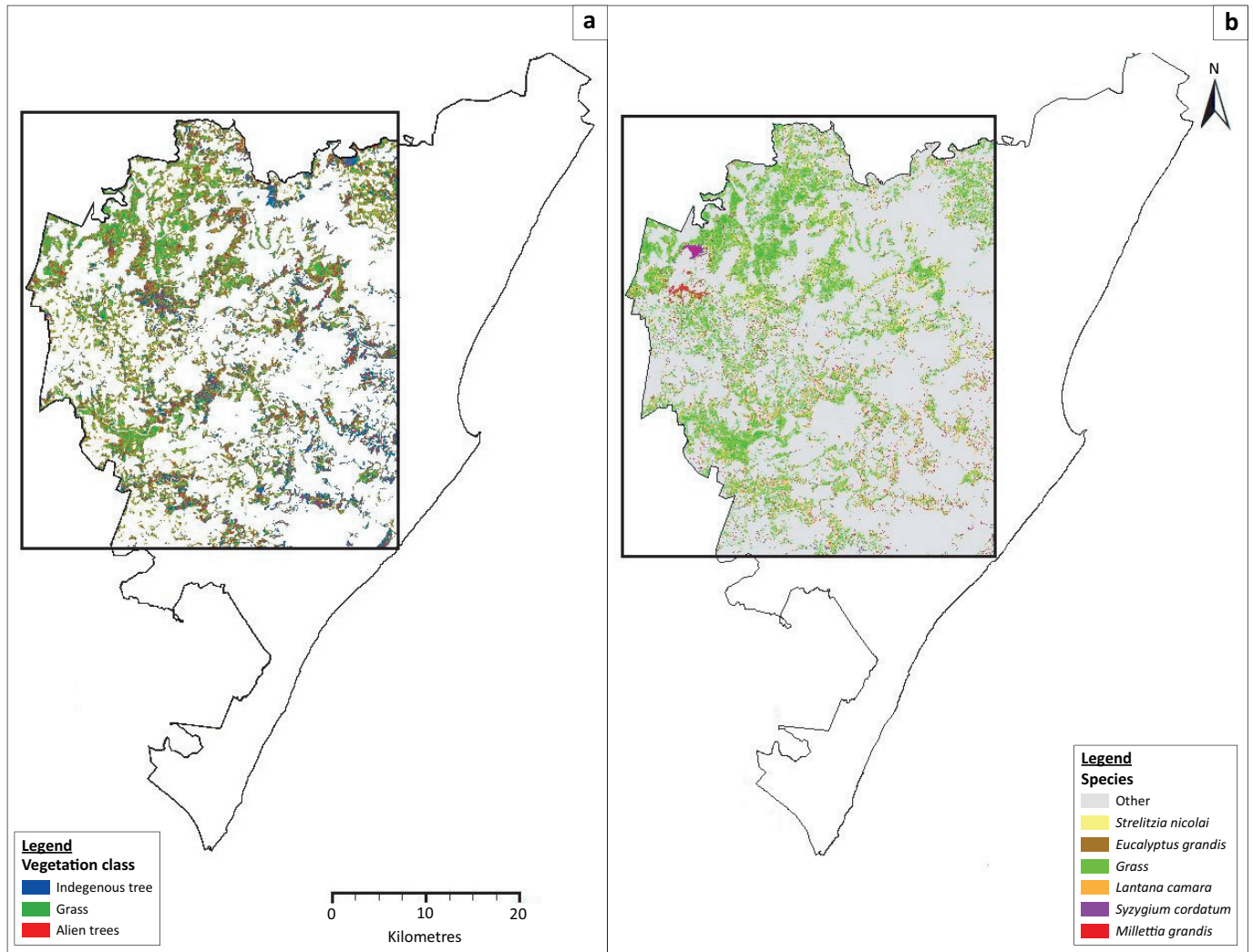
Source: Authors' own work

(a), the distribution of three indigenous (*Syzygium cordatum, Strelitzia nicolai* and *Millettia grandis*); (b), and two alien (*Lantana camara* and *Eucalyptus grandis*) plant species.

FIGURE 2: Indigenous and alien trees in the eThekwini Municipality using Random Forest
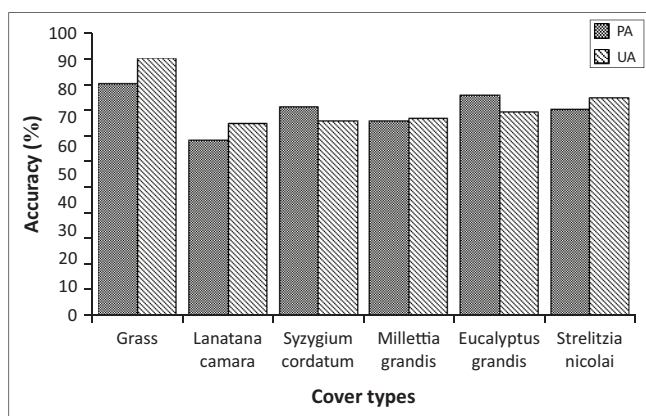
Source: Authors' own work

FIGURE 3: User and producer accuracies obtained from the Random Forest classification algorithm for dominant cover types.

According to Gilmore *et al.* (2008), vegetation spectral characteristics and species separability is determined by leaf pigmentation, water content, leaf size and leaf structure, amongst others. The additional red-edge band available on the RapidEye sensor has been shown to be sensitive to subtle variability in these characteristics, hence valuable for discriminating plant species (Cho *et al.* 2012). The spectral configuration and improved spatial resolution that characterise new generation sensors like RapidEye offer new opportunities in land cover mapping (Govender *et al.* 2008). Specifically, the increased number of bands in these sensors facilitates the discrimination of surface objects with subtle reflectance variation. Novack *et al.* (2011), for instance, showed an improved classification accuracy using new generation sensors with additional strategically positioned bands in comparison to the Quickbird-2 sensor (with four traditional bands) on an urban landscape. The red-edge band is particularly valuable in discriminating vegetation classes and species (Safri, Salleh & Ghiyamat 2006). According to Cho and Skidmore (2006), the red-edge position is similar amongst different vegetation species and is sensitive to variation in chlorophyll content and internal leaf structure. As plant chlorophyll content and internal leaf structure vary between species, the band has shown great promise for inter-species and intra-species discrimination, determination of biomass quantity, stages of crop development and plant health (Gitelson & Merzlyak 1994). In this study, over 86%

overall accuracy was achieved for the major vegetation categories (i.e. grass, indigenous trees and alien vegetation) whilst over 68% overall accuracy was achieved for the dominant species. Although the influence of the red-edge on classification accuracy was not tested, the reliable classification accuracy achieved is consistent with Adelabu *et al.* (2013) who established a 78%–80.25% and 85%–88.75% accuracy when the red-edge band was excluded and included in the classification, respectively.

A suite of traditional classification algorithms has been developed for remotely sensed imagery. However, amongst the existing algorithms, the maximum likelihood and minimum distance to mean approaches have been popular because of their ability to generate acceptable classification results, simplicity and availability in popular remote sensing software (Lu & Weng 2005; Song, Duan & Jiang 2012). However, Kavzoglu and Mather (2003) note that the performance of these traditional algorithms is constrained by their data distributional assumptions and data input restriction. Efforts for improved classification accuracy on remotely sensed imagery has seen the development of advanced classification techniques that include support vector machines, neural networks and decision trees (Lu & Weng 2007). Recently, a number of studies (e.g. Chan *et al.* 2012; Peters *et al.* 2011; Watts *et al.* 2009) have demonstrated the superiority of the RF approach. In this study, results obtained using the RF algorithm demonstrated good classification accuracies for individual vegetation classes. A number of studies (e.g. Dixon & Candade 2008; Pal & Mather 2004; Song *et al.* 2012) have established that the new, advanced and robust non-parametric algorithms like RF are superior to the commonly used parametric approaches like maximum likelihood. In this study, 93%, 88% and 82% overall accuracy was established for grass, indigenous and alien vegetation, respectively. The high classification accuracy achieved in the study is consistent with Waske and Braun (2009), who attributed RF's high classification approach to its non-parametric approach on handling remotely sensed data. These findings are also consistent with Naidoo *et al.* (2012), who achieved 88% accuracy in classifying woody species within the Kruger National Park's savannah vegetation using the algorithm. Watts *et al.* (2009) note that, unlike existing statistical approaches, RF has no distributional assumptions on input data set, is not affected by over-fitting, has the ability to handle unbalanced data sets and is computationally efficient. Therefore, non-parametric approaches like RF are a viable option for generating robust classifications from complex image brightness, and are suitable for highly multifarious landscapes (Ghimire, Rogan & Miller 2010). The high classification accuracy achieved in grass can be attributed to its greater spectral variation from woody species, thus reducing the probability of misclassification (Abdel-Rahman *et al.* 2014a).

Results in this study showed high (86% and 88%) individual accuracies for alien and indigenous classes, respectively. Spectrally, the two classes are relatively similar, particularly their reflectance at the green and NIR sections of the electromagnetic spectrum. However, as mentioned, the higher classification accuracy can be attributed to the sensor's higher spatial and spectral resolution, which includes the red-edge band that facilitates the discrimination of species (Oldeland *et al.* 2010). Relatively good accuracies, ranging between 68% and 95%, were achieved for dominant vegetation species (Figure 3). In land cover mapping, a 70% classification accuracy is often considered ideal (Thomlinson, Bolstad & Cohen 1999). In our study, we acknowledge that the 68% classification accuracy for *Lantana camara* was below the required threshold. We attribute this to its indistinct spectral characteristics that varies with growth stages, causing spectral confusion with other species. Furthermore, *Lantana camara* is often found in heterogeneity with other vegetation types, causing confusion. Nevertheless, we consider the 68% classification accuracy valuable for initial screening for areas with *Lantana camara* invasion. Generally, these findings are consistent with Adelabu *et al.* (2013), who discriminated five indigenous and exotic tree species with over 70% overall classification accuracy and 10% allocation and 3% quantity disagreement scores.

Larger patches of the indigenous and alien species are predominantly found in the western part of eThekwini Municipality. Furthermore, this study shows that there is significant indigenous and alien plant species cover within the KZN SS. This study therefore offers valuable insight into the distribution of woody and alien vegetation within the grassland, valuable for management and optimisation of the grassland.

The aim of this study was to map woody vegetation in the KZN SS using satellite remote sensing. The results of the study have shown that:

- RF was successful in discriminating between the two types of woody cover, with an overall accuracy and individual accuracy above 80%
- RF was successful in differentiating between five tree species (both indigenous and alien) with individual accuracy values above 68%
- low quantity disagreement and allocation disagreement scores asserted the robustness of the model.

Overall, the results of this study have shown the importance of new generation sensors in mapping and discriminating the distribution of trees on a landscape scale. Specifically, this study provides an understanding of alien and woody vegetation encroachment within the KZN SS, valuable for optimising relevant conservation management plans for restoration or maintenance of the grassland's ecological integrity.

Acknowledgements
Competing interests

Authors' contributions

J.O. conceptualised the research, wrote the original draft and prepared the manuscript for submission. O.M. conceptualised the research, advised on data collection, analysed the data and wrote the original and final drafts. M.R. conceptualised the research, advised on data analysis and wrote the original manuscript draft. N.H. wrote the original manuscript and collected and analysed the data. All authors provided equal input to the manuscript.

References

Abdel-Rahman, E.M., Mutanga, O., Adam, E. & Ismail, R., 2014a, 'Detecting *Sirex noctilio* grey-attacked and lightning-struck pine trees using airborne hyperspectral data, random forest and support vector machines classifiers', *ISPRS Journal of Photogrammetry and Remote Sensing* 88, 48–59. http://dx.doi.org/10.1016/j.isprsjprs.2013.11.013

Abdel-Rahman, E.M., Mutanga, O., Odindi, J., Adam, E., Odindo, A., & Ismail, R., 2014b, 'A comparison of partial least squares (PLS) and sparse PLS regressions for predicting yield of Swiss chard grown under different irrigation water sources using hyperspectral data', *Computers and Electronics in Agriculture* 106, 11–19. http://dx.doi.org/10.1016/j.compag.2014.05.001

Adam, E., Mutanga, O., Odindi, J. & Abdel-Rahman, E., 2014, 'Land-use/cover classification in a heterogeneous coastal landscape using RapidEye imagery: Evaluating the performance of random forest and support vector machines classifiers', *International Journal of Remote Sensing* 35, 3440–3458. http://dx.doi.org/10.1080/01431161.2014.903435

Adelabu, S., Mutanga, O. & Adam, E., 2014, 'Evaluating the impact of red-edge band from RapidEye image for classifying insect defoliation levels', *ISPRS Journal of Photogrammetry and Remote Sensing* 95, 34–41. http://dx.doi.org/10.1016/j.isprsjprs.2014.05.013

Adelabu, S., Mutang, O. & Adam, E., 2015, 'Testing the reliability and stability of the internal accuracy assessment of random forest using different validation methods', *Geocarto International* 30, 810–821. http://dx.doi.org/10.1080/10106049.2014.997303

Adelabu, S., Mutanga, O., Adam, E. & Cho, A.C., 2013, 'Exploiting machine learning algorithms for tree species classification in a semiarid woodland using RapidEye image', *Journal of Applied Remote Sensing* 7, 073480. http://dx.doi.org/10.1117/1.JRS.7.073480

Belluco, E., Camuffo, M. & Ferrari, S., 2006, 'Mapping salt-marsh vegetation by multispectral and hyperspectral remote sensing', *Remote Sensing of Environment* 105, 54–67. http://dx.doi.org/10.1016/j.rse.2006.06.006

Breiman, L., 2001, 'Random forests', *Machine Learning* 45, 5–32. http://dx.doi.org/10.1023/A:1010933404324

Briggs, J.M., Schaafsma, H. & Trekov, D., 2007, 'Woody vegetation expansion in a desert grassland: Prehistoric human impact?', *Journal of Arid Environments*, 69, 458–472. http://dx.doi.org/10.1016/j.jaridenv.2006.10.012

Carlier, L., Rotar, I., Vlahova, M. & Vidican, R., 2009, 'Importance and functions of grasslands', *Notulae Botanicae Horti Agrobotanici Cluj-Napoca* 37(1), 25–30.

Chan, J.C., Beckers, P., Spanhove, T. & Borre, J.V., 2012, 'An evaluation of ensemble of classifiers for mapping Natura 2000 heatland in Belgium using spaceborne angular hyperspectral (CHRIS/Proba) imagery', *International Journal of Applied Earth Observation and Geo-information* 18, 13–22. http://dx.doi.org/10.1016/j.jag.2012.01.002

Chapman, D.S., Bonn, A., Kunin, W.E. & Cornell, S.J., 2009, 'Random forest characterization of upland vegetation and management burning from aerial imagery', *Journal of Biogeography* 37, 37–46. http://dx.doi.org/10.1111/j.1365-2699.2009.02186.x

Cho, M.A., Rouget, M., Asner, G., Naidoo, L., van Aardt, J., Ramoelo, A., *et al.*, 2012, 'Mapping tree species composition in South Africa savannas using and integrated airborne spectral and Lidar system', *Remote Sensing of Environment* 125, 214–226. http://dx.doi.org/10.1016/j.rse.2012.07.010

Cho, M.A. & Skidmore, A., 2006, 'A new technique for extracting the red edge position from hyperspectral data: The linear extrapolation method', *Remote Sensing of Environment* 101, 181–193. http://dx.doi.org/10.1016/j.rse.2005.12.011

Congalton, R., 1991, 'A review of assessing the accuracy of classifications of remotely sensed data', *Remote Sensing of Environment* 37(1), 35–46. http://dx.doi.org/10.1016/0034-4257(91)90048-B

Department of Rural Development and Land Reform, 2009, *Development of a methodology for national land use mapping for chief directorate: Geo-spatial information*, Chief Directorate, Geo-spatial Information, Mowbray.

Dixon, B. & Candade, N., 2008, 'Multispectral landuse classification using neural networks and support vector machine: One or the other, or both?', *International Journal of Remote Sensing* 29, 1185–1206. http://dx.doi.org/10.1080/01431160701294661

Eitel, J.U.H., Long, D.S., Gessler, P.E. & Smith, A.M.S., 2007, 'Using in-situ measurements to evaluate the new RapidEye satellite series for prediction of wheat nitrogen status', *International Journal of Remote Sensing* 28, 4183–4190. http://dx.doi.org/10.1080/01431160701422213

Environmental Systems Research Institute, 2016, *ArcGIS 10.3.1*, Environmental Systems Research Institute (ESRI), Redlands, CA.

eThekwini Municipality, 2012, 'State of Biodiversity Report: 2011/2012', Environmental Planning and Climate Protection Department, eThekwini Municipality, Durban, viewed 8 August 2016, from www.durban.gov.za/City_Services/development_planning_manangement/environmental_plan_climate_protection/publications/Documents/State_of_Biodiversity_Report_2011_2012.pdf

Fraser, R.H., Abuelgasim, A. & Latifovic, R., 2005, 'A method for detecting large-scale forest cover change using coarse spatial resolution imagery', *Remote Sensing of Environment* 95, 414–427. http://dx.doi.org/10.1016/j.rse.2004.12.014

Ghimire, B., Rogan, J. & Miller, J., 2010, 'Contextual land cover classification: Incorporating spatial dependence in land cover classification models using random forest and Getis statistic', *Remote Sensing Letters* 1, 45–54. http://dx.doi.org/10.1080/01431160903252327

Gilmore, M.S., Wilson, E.H., Barret, N., Civco, D.L., Prisloe, S., Hurd, J.D., *et al.*, 2008, 'Integrating multi-temporal and structural information to map wetland vegetation in a lower Connecticut River tidal marsh', *Remote Sensing of Environment* 112, 4048–4060. http://dx.doi.org/10.1016/j.rse.2008.05.020

Gitelson, A. & Merzlyak, M., 1994, 'Quantitative estimation of chlorophyll-a using reflectance spectra: Experiments with autumn chesnut and maple leaves', *Journal of Photochemistry and PhotobiologyB: Biology* 22, 247–252. http://dx.doi.org/10.1016/1011-1344(93)06963-4

Govender, M., Chetty, K., Naiken, V. & Bulcock H., 2008, 'A comparison of satellite hyperspectral and multispectral remote sensing imagery for improved classification and mapping of vegetation', *Water SA* 34, 147–154.

Guo, L., Chehata, N., Mallet, C. & Boukir, S., 2011, 'Relevance of airborne lidar and multispectral image data for urban scene classification using random forest', *ISPRS Journal of Photogrammetry and Remote Sensing* 66, 56–66. http://dx.doi.org/10.1016/j.isprsjprs.2010.08.007

Ham, J., Chen, Y., Crawford, M.M. & Gosh, J., 2005, 'Investigating the random forest framework for classification of hyperspectral data', *IEEE Transactions on Geoscience and Remote Sensing* 43, 492–501. http://dx.doi.org/10.1109/TGRS.2004.842481

Henderson, B.L., Bui, E.N., Moran, C.J. & Simon, D.A.P., 2005, 'Australia-wide predictions of soil properties using decision trees', *Geoderma* 124, 383–398. http://dx.doi.org/10.1016/j.geoderma.2004.06.007

Ismail, R. & Mutanga, O., 2010, 'A comparison of egression tree ensemble: Predicting *Sirex noctilio* induced water stress in *Pinus patula* forest of KwaZulu-Natal, South Africa', *International Journal of Applied Earth Observation and Geoinformation* 12, 45–51. http://dx.doi.org/10.1016/j.jag.2009.09.004

Jewitt, D., 2011, 'Conservation targets and status for vegetation types in KZN', Unpublished internal report, Biodiversity Conservation Planning Division, Ezemvelo KZN Wildlife, viewed 8 August 2016, from http://www.bgis.sanbi.org/KZN/KZN_vegetationtypes_conservationtargetstatusreport.pdf

Kavzoglu, T. & Mather, P., 2003, 'The use of backpropagating artificial neural networks in land cover classification', *International Journal of Remote Sensing* 24, 4907–4938. http://dx.doi.org/10.1080/0143116031000114851

Kiala, Z., Odindi, J., Mutanga, O. & Peerbhay K., 2016, 'Comparison of partial least squares and support vector regressions for predicting leaf area index on a tropical grassland using hyperspectral data', *Journal of Applied Remote Sensing* 10, 036015. http://dx.doi.org/10.1117/1.JRS.10.036015

Lalla R., 2014, 'South Africa – A global player in the battle against alien plant invasions', *South African Journal of Science* 110, 7–8. http://dx.doi.org/10.1590/sajs.2014/a0073

Lawrence, R.L., Wood, S.D. & Sheley, R.L., 2006, 'Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (randomForest)', *Remote Sensing of Environment* 100, 356–362. http://dx.doi.org/10.1016/j.rse.2005.10.014

Lin, X., Sun, L., Li, Y., Guo, Z., Li, Y., Zhong, K., *et al.*, 2010, 'A random forest of combined features in the classification of cut tobacco based on gas chromatography fingerprinting', *Talanta* 82, 1571–1575. http://dx.doi.org/10.1016/j.talanta.2010.07.053

Loosevelt, L., Peters, J., Skriver, H., Lievens, H., van Coillie, F., De Baets, B., *et al.*, 2012, 'Random forest as a tool for estimating uncertainity at pixel-level in SAR image classification', *International Journal of Applied Earth Observation and Geoinformation* 19,173–184. http://dx.doi.org/10.1016/j.jag.2012.05.011

Lu, D. & Weng, Q., 2004, 'Spectral mixture analysis of the urban landscape in Indianapolis with Landsat ETM+ imagery', *Photogrammetric Engineering and Remote Sensing* 70, 1053–1062. http://dx.doi.org/10.14358/PERS.70.9.1053

Lu, D. & Weng, Q., 2005, 'Urban classification using full spectral information of Landsat ETM+ imagery in Marion County, Indiana', *Photogrammetric Engineering and Remote Sensing* 71, 1275–1284. http://dx.doi.org/10.14358/PERS.71.11.1275

Lu, D. & Weng, Q., 2007, 'A survey of image classification methods and techniques for improving classification performance', *International Journal of Remote Sensing* 28, 823–870. http://dx.doi.org/10.1080/01431160600746456

Mansour, K., Mutanga, O. & Everson, T., 2012, 'Spectral discrimination of increaser species as an indicator of rangeland degradation using field spectrometry', *Journal of Spatial Science* 58(1), 101–117. http://dx.doi.org/10.1080/14498596.2012.759090

Martindale, G., 2007, 'Influence of livestock grazing on plant diversity of Highland Sourveld grassland in KwaZulu-Natal', PhD thesis, School of Animal, Plant and Environmental Sciences, University of Witwatersrand, Johannesburg, South Africa.

Matsika, R., 2007, 'Land-cover change: Threats to the grassland biome of South Africa', MSc thesis, School of Animal, Plant and Environmental Sciences, University of the Witwatersrand, Johannesburg, South Africa.

Melgani, F. & Bruzzone, L., 2004, 'Classification of hyperspectral remote sensing images with support vector machines', *IEEE Transactions on Geoscience and Remote Sensing* 42, 1778–1790. http://dx.doi.org/10.1109/TGRS.2004.831865

Mountrakis, G., Im, J. & Ogole, C., 2011, 'Support vector machines in remote sensing: A review', *ISPRS Journal of Photogrammetry and Remote Sensing* 66, 247–259. http://dx.doi.org/10.1016/j.isprsjprs.2010.11.001

Mucina, L. & Rutherford, M.C., 2006, *The vegetation of South Africa, Lesotho and Swaziland*, South African National Biodiversity Institute, Pretoria.

Munyati, C., Shaker, P. & Phasha, M.G., 2011. 'Using remotely sensed imagery to monitor savanna rangeland deterioration through woody plant proliferation: A case study from communal and biodiversity conservation rangeland sites in Mokopane, South Africa', *Environmental Monitoring Assessment* 176, 293–311. http://dx.doi.org/10.1007/s10661-010-1583-4

Naidoo, L., Cho, M.A., Rouget, M. & Asner, G., 2012, 'Classification of savanna tree species, in the Greater Kruger National Park region, by integrating hyperspectral and LiDAR data in a Random Forest data mining environment', *ISPRS Journal of Photogrammetry and Remote Sensing* 69, 167–179. http://dx.doi.org/10.1016/j.isprsjprs.2012.03.005

Novack, T., Esch, T., Kux, H. & Stilla, U., 2011, 'Machine Learning Comparison between WorldView-2 and QuickBird-2-Simulated Imagery Regarding Object-Based Urban Land Cover Classification', *Remote Sensing* 3, 2263–2282. http://dx.doi.org/10.3390/rs3102263

Odindi, J., Mhangara, P. & Kakembo, V., 2012, 'Remote sensing land-cover change in Port Elizabeth during South Africa's democratic transition', *South African Journal of Science* 108(5/6), 886. http://dx.doi.org/10.4102/sajs.v108i5/6.886

Ojoyi, M., Mutanga, O., Odindi, J. & Abdel-Rahman, E., (2016), 'Application of topo-edaphic factors and remotely sensed vegetation indices to enhance biomass estimation in a heterogeneous landscape in the Eastern Arc Mountains of Tanzania', *Geocarto International* 31, 1–21. http://dx.doi.org/10.1080/10106049.2015.1041557

Oldeland, J., Dorigo, W., Wesuls, D. & Jürgens, N., 2010, 'Mapping bush encroaching species by seasonal differences in hyperspectral imagery', *Remote Sensing* 2, 1416–1438. http://dx.doi.org/10.3390/rs2061416

Pal, M., 2005, 'Random forest classifier for remote sensing classification', *International Journal of Remote Sensing* 26, 217–222. http://dx.doi.org/10.1080/01431160412331269698

Pal, M. & Mather, P., 2004, 'Assessment of the effectiveness of support vector machines for hyperspectral data', *Future Generation Computer Systems* 20, 1215–1225. http://dx.doi.org/10.1016/j.future.2003.11.011

Peerbhay, K.Y., Mutanga, O. & Ismail, R. 2013, 'Commercial tree species discrimination using airbone AISA Eagle hyperspectral imagery and partial least squares discriminant analysis (PLS-DA) in KwaZulu-Natal, South Africa', *ISPRS Journal of Photogrammetry and Remote Sensing* 79, 19–28. http://dx.doi.org/10.1016/j.isprsjprs.2013.01.013

Peters, J., De Baets, B., Verhoest, N.E.C., Samson, R., Degroeve, S., De Becker, P., *et al.*, 2007, 'Random forest as a tool for ecological distribution modeling', *Ecological Modelling*, 207 304–318. http://dx.doi.org/10.1016/j.ecolmodel.2007.05.011

Peters, J., van Coillie, F., Westra, T. & de Wulf, R., 2011, 'Synergy of very high resolution optical and radar data for object-based olive grove mapping', *International Journal of Geographical Information Science* 25, 971–989. http://dx.doi.org/10.1080/13658816.2010.515946

Pontius, R.G. & Millones, M., 2011, 'Death to Kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment', *International Journal of Remote Sensing* 32, 4407–4429. http://dx.doi.org/10.1080/01431161.2011.552923

Roberts, D., Boon, R., Diederichs, N., Douwes, E., Govender, N., Mcinnes, A., *et al.*, 2012, 'Exploring ecosystem-based adaptation in Durban, South Africa: "Learning-bydoing" at the local government coal face', *Environment and Urbanization* 24, 167–195. http://dx.doi.org/10.1177/0956247811431412

Rogan, J. & Chen, D.M., 2004, 'Remote sensing technology for mapping and monitoring land-cover land-use change', *Progress in Planning* 61, 301–325. http://dx.doi.org/10.1016/S0305-9006(03)00066-7

Rogan, J., Franklin, J., Stow, D., Miller, J., Woodcock, C. & Roberts, D., 2008, 'Mapping land-cover modifications over large areas: A comparison of machine learning algorithms', *Remote Sensing of Environment* 112, 2272–2283. http://dx.doi.org/10.1016/j.rse.2007.10.004

Rutherford, M.C., Mucina, L., Lötter, M.C., Bredencamp, G.J., Smit, J.H.L., Scot-Shaw, R., *et al.*, 2006. 'Savanna biome', in L. Mucina & M.C. Rutherford (eds.), *The vegetation of South Africa, Lesotho and Swaziland*, pp. 439–539, South African National Biodiversity Institute, Pretoria.

Safri, H.Z.M., Salleh, M.A.M. & Ghiyamat, A., 2006, 'Hypespectral remote sensing of vegetation using red-edge position techniques', *American Journal of Applied Sciences* 3, 1864–1871. http://dx.doi.org/10.3844/ajassp.2006.1864.1871

Schuster, C.M., Forster, M. & Kleinschmit, B., 2012, 'Testing the red edge channel for improving land-use classification based on high-resolution multi-spectral satellite data', *International Journal of Remote Sensing* 33, 5583–5599. http://dx.doi.org/10.1080/01431161.2012.666812

Sesnie, S.E., Finegan, B., Gessler, P.E., Thessler, S., Bendana, Z.R. & Smith, A.M.S., 2010, 'The multispectral separability of Costa Rican rainforest types with support vector machines and random forest decision trees', *International Journal of Remote Sensing* 31, 2885–2909. http://dx.doi.org/10.1080/01431160903140803

Shekede, M.D., Murwira, A. & Masocha, M., 2015, 'Wavelet-based detection of bush encroachment in a savanna using multi-temporal aerial photographs and satellite imagery', *International Journal of Applied Earth Observation and Geoinformation* 35, 209–216. http://dx.doi.org/10.1016/j.jag.2014.08.019

Song, X., Duan, Z. & Jiang, X., 2012, 'Comparison of artificial neural networks and support vector machine classifiers for land cover classification in northern China using SPOT-5 HRG image', *International Journal of Remote Sensing* 33, 3301–3320. http://dx.doi.org/10.1080/01431161.2011.568531

Thenkabail, P., Enclona, E., Ashton, M. & Van Der Meer, B., 2004, 'Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications', *Remote Sensing of Environment* 91, 354–376. http://dx.doi.org/10.1016/j.rse.2004.03.013

Thomlinson, J.R., Bolstad, P.V. & Cohen, W.B., 1999, 'Coordinating methodologies for scaling landcover classifications from site-specific to global: Steps toward validating global map products', *Remote Sensing of Environment* 70, 16–28. http://dx.doi.org/10.1016/S0034-4257(99)00055-3

Tigges, J., Lakes, T. & Hostert, P., 2013, 'Urban vegetation classification: Benefits of multi-temporal RapidEye satellite data', *Remote Sensing of Environment* 136, 66–75. http://dx.doi.org/10.1016/j.rse.2013.05.001

Van Auken, O.W., 2009, 'Causes and consequences of woody plant encroachment into western North American grasslands', *Journal of Environmental Management* 90, 2931–2942. http://dx.doi.org/10.1016/j.jenvman.2009.04.023

Waske, B. & Braun, M., 2009, 'Classifier ensemble for land cover mapping using multitemporal SAR imagery', *ISPRS Journal of Photogrammetric and Remote Sensing* 65, 450–457. http://dx.doi.org/10.1016/j.isprsjprs.2009.01.003

Watanachaturaporn, P., Arora, M.K. & Varshney, P.K., 2008, 'Multi-source classification using support vector machine: An empirical comparison with decision tree and neural network classifiers', *Photogrammetric Engineering and Remote Sensing* 74, 239–246. http://dx.doi.org/10.14358/PERS.74.2.239

Watts, J.D., Lawrence, R.L., Miller, P.R. & Montange, C., 2009, 'Monitoring cropland practices for carbon sequestration purposes in north central Montana by Landsat remote sensing', *Remote Sensing of Environment* 113, 1843–1852. http://dx.doi.org/10.1016/j.rse.2009.04.015

World Resources Institute, 2000, 'Forests, grasslands and drylands', viewed 5 August 2016, from http://earthtrends.wri.org/text/forests-grasslands-drylands/map-229.htm

Xie, Y., Sha, Z. & Yu, M., 2008, 'Remote sensing imagery in vegetation mapping: A review', *Journal of Plant Ecology* 1, 9–23. http://dx.doi.org/10.1093/jpe/rtm005

Yuan, F., Bauer, M.E., Heinert, N.J. & Holden, G.R., 2005, 'Multi-level land cover mapping of twin cities (Minnesota) metropolitan area with multi-seasonal Landsat TM/ETM+ data', *Geocarto International* 20, 5–13. http://dx.doi.org/10.1080/10106040508542340