Geocarto International
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tgei20

Multispectral remote sensing for mapping grassland degradation using the key indicators of grass species and edaphic factors

Khalid Mansour[ab], Onisimo Mutanga[a], Elhadi Adam[ac] & Elfatih M. Abdel-Rahman[ad]

[a] School of Agricultural, Environmental and Earth Sciences, University of KwaZulu-Natal (UKZN), Pietermaritzburg, South Africa

[b] Department of Geography, University of Al-Fashir, Al-Fashir, Sudan

[c] School of Geography, Archaeology and Environmental Studies, University of the Witwatersrand, Johannesburg, South Africa

[d] Faculty of Agriculture, Department of Agronomy, University of Khartoum, Khartoum North, Sudan

Accepted author version posted online: 10 Jun 2015.Published online: 01 Jul 2015.

Geocarto International, 2015
http://dx.doi.org/10.1080/10106049.2015.1059898

Multispectral remote sensing for mapping grassland degradation using the key indicators of grass species and edaphic factors

Khalid Mansour[a,b], Onisimo Mutanga[a], Elhadi Adam[a,c] and Elfatih M. Abdel-Rahman[a,d]*

[a]*School of Agricultural, Environmental and Earth Sciences, University of KwaZulu-Natal (UKZN), Pietermaritzburg, South Africa;* [b]*Department of Geography, University of Al-Fashir, Al-Fashir, Sudan;* [c]*School of Geography, Archaeology and Environmental Studies, University of the Witwatersrand, Johannesburg, South Africa;* [d]*Faculty of Agriculture, Department of Agronomy, University of Khartoum, Khartoum North, Sudan*

(*Received 9 September 2014; accepted 4 June 2015*)

Land degradation is believed to be one of the most severe and widespread environmental problems. In South Africa, large areas of land have been identified as degraded, as shown by the lower vegetation cover. One of the major causes of grassland degradation is change in plant species composition that leads to presence of unpalatable grass species. Some grass species have been successfully used as indicators of different levels of grassland degradation in the country. This paper, therefore explores the possibility of mapping grassland degradation in Cathedral Peak, South Africa, using indicators of grass species and edaphic factors. Multispectral SPOT 5 data were used to produce a grassland degradation map based on the spatial distribution of decreaser (*Themeda triandra*) and increaser (*Hyparrhenia hirta*) species. To improve mapping accuracy, soil samples were collected from each species site and analysed for nutrient content. A *t*-test and machine learning random forest classification algorithm were applied for variable selection and classification using SPOT 5 data and edaphic variables. Results indicated that the decreaser and increaser grass species can be mapped with modest accuracy using SPOT 5 data (overall accuracy of 75.30%, quantity disagreement = 2 and allocation disagreement = 23). The classification accuracy was improved to 88.60%, 1 and 11 for overall accuracy, quantity and allocation disagreements, respectively, when SPOT 5 bands and edaphic factors were combined. The study demonstrated that an approach based on the integration of multispectral data and edaphic variables, which increased the overall classification accuracy by about 13%, is a suitable when adopting remote sensing to monitor grassland degradation.

Keywords: grassland degradation; key indicators; edaphic factors; SPOT 5 XS; random forest

1. Introduction

Land degradation is believed to be one of the most severe and widespread environmental problems in South Africa (Hoffman & Todd 2000; Wessels et al. 2004). In South Africa, a large area of land (4.8%) has been identified as degraded because of low vegetation cover compared to the surrounding (Thompson 1996; Wessels et al. 2004). The value of grassland species can be considered in terms of their types and communities and productivity, particularly their palatability to livestock (Tainton 1999). One of the major causes

*Corresponding author. Email: MutangaO@ukzn.ac.za

of grassland degradation is change in plant species composition that leads to presence of unpalatable grass species. In South Africa, certain grass species have been successfully used as indicators of different levels of grassland degradation (Mansour & Mutanga 2012; Mansour et al. 2012). This is because plant species are well adapted to specific growth conditions, and their numbers will reduce or increase dramatically if these conditions change (van Oudtshoorn 1999; van den Berg & Zeng 2006). Based on changes in their relative abundance when overgrazed or undergrazed, grassland species can be grouped into two categories, increasers and decreasers (Dobarro et al. 2010). Increaser species increase their relative abundance through overgrazing or under-utilization, and are therefore indicators of low land productivity (van Oudtshoorn 1999; Dobarro et al. 2010), while decreaser species decrease their relative abundance when rangeland condition deteriorates through over or under-utilization (van Oudtshoorn 1999; Dobarro et al. 2010). Despite the relative importance of decreaser species for grazing in South African rangelands, overgrazing and intense agricultural activities have threatened the existence of decreaser species (Tainton 1999). Decreaser and increaser communities constitute a convenient means to monitor deterioration or improvement of land productivity (Tainton 1999). The spatial distribution of decreaser and increaser species also need to be determined to sustain natural resource management in general and to provide valuable information for understanding grassland productivity in particular.

Mapping grassland degradation requires accurate information about the spatial distribution of decreaser and increaser grass species. Using traditional survey methods such as local expert knowledge and field observation, these methods provide significantly better results for mapping grassland degradation over small geographic areas. However, these methods require intensive fieldwork, which is costly and time-consuming (Muchoney & Haack 1994). Grassland degradation monitoring, detection, and mapping using remotely sensed data, on the other hand, have attracted scientific attention. Satellite images of different spatial and spectral resolutions are cost-effective, with timely and relatively accurate information (Lees & Ritman 1991; Ustin et al. 2009) that are useful for monitoring grassland degradation. The use of finer spectral resolution hyperspectral data remains important for vegetation species discrimination because of their detailed spectral information (Kumar et al. 2001; Thenkabail et al. 2004). These detailed features enable the identification of spectral attributes for discriminating and monitoring rangeland degradation at species levels (Mansour et al. 2012). However, compared to multispectral data, hyperspectral data are expensive and require skilled processing knowledge (Underwood et al. 2003; Schmidtlein & Sassin 2004; Lawrence et al. 2006). Multispectral data, on the other hand, are relatively easily available and inexpensive, and do not require complex pre-processing and processing techniques. Due to these advantages, use of multispectral data is therefore for accurate and up-to-date mapping of grassland species over large areas.

Multispectral data have been used for several decades to map vegetation communities in different ecosystems (e.g. Rignot et al. 1997; Harvey & Hill 2001; Chastain et al. 2008; Martínez-López et al. 2014). Although results in these studies have been promising, mapping vegetation in complex environments at species level using multispectral imageries (e.g. Landsat TM and SPOT) has often been impeded by complex spatial and spectral patterns of species in heterogeneous vegetation canopies. Recently, researchers have advocated for the use of multispectral data from the newly launched sensors (e.g. WorldView-2) to map vegetation at species level. The finer spatial resolution and strategically located bands (e.g. red edge) of the new sensors allow vegetation species mapping at relatively higher accuracy (Cho et al. 2012; Mansour & Mutanga 2012).

Mansour and Mutanga (2012), for example, have utilized WorldView-2 data to map grass species to assess grassland degradation in South African ecosystem. Whereas their study resulted in a successful discrimination (overall accuracy = 90%, kappa = 0.87) between decreaser and increaser grass species, they regarded the cost of WorldView-2 imagery as relatively high (ca 50 US$/km^{-2}) for an operational grass species discrimination. Readily accessible satellite-based data of moderately high spectral and spatial resolutions would offer a better opportunity for an operational application for large-scale grassland degradation assessment. It is worth noting that South African National Space Agency (SANSA) provides a country-wide SPOT 5 data for free (SANSA 2012). SPOT 5 provides an ideal balance between small pixel size (10 × 10 m) and wider-area coverage (60 × 60 km) with four spectral bands, suitable for monitoring grassland degradation.

Additionally, in order to improve the classification accuracy of mapping vegetation at species level, researchers have integrated some topographic variables together with multispectral data. Integration of topographic variables (e.g. slope, aspect, and elevation) and Landsat 7 ETM+ data was used to map nine vegetation groups with an increase in overall accuracy of 19.6% (DomaÇ & Süzen 2006). In another study, degraded lands were mapped using Landsat TM and ETM+ with relatively low overall accuracies of 53 and 72%, respectively. The overall classification accuracy was further improved by 10% (ETM+) and 14% (TM) when slope and aspect variables were integrated with Landsat data (Liberti et al. 2009). However, more research is needed to determine the value of incorporating multispectral data of relatively high spatial resolution (e.g. SPOT 5) with ancillary data such as edaphic variables to improve the classification accuracy of species-level vegetation mapping. Since our interest is in monitoring grassland degradation, we tested the hypothesis that edaphic factors influence the growth and distribution of decreaser and increaser grass species.

Soil macro and micronutrients (total N, available P, soluble K, Ca and Mn, Mg and Zn, respectively) play an important role in the plant bio-physiological processes like photosynthesis and respiration (Marschner 1993). The efficiencies and rates of plant bio-physiological processes determine the rate of vegetation growth and development and other vegetation structural (phenotypic) characteristics (Chapin 1980; Marschner & Rimmington 1988). The differences in these vegetation physiological and morphological properties could therefore result in different vegetation spectral features (Lillesand & Kiefer 2001; Kumar et al. 2001). Moreover, grassland degradation could be a result of degraded soil conditions (Dlamini et al. 2014). It is known that during the growing season, soil spectral properties are masked out by vegetation cover. The integration of some soil variables together with remotely sensed data in monitoring grassland degradation could hence provide an opportunity to indirectly determine soil degradation. The abundance of decreaser grass species, for example, could be due to soil degradation. The objective of the present study was to test the utility of the freely available SPOT 5 multispectral data and some edaphic variables in monitoring grassland degradation in a South African ecosystem based on the spatial distribution of decreaser and increaser grass species.

2. Methodology

2.1. Study area

Cathedral Peak (29°00′ E to 29°30′ E and 28°45′ S to 29°15′ S) is located in the northern part of the Natal Drakensberg Park, KwaZulu-Natal province, South Africa

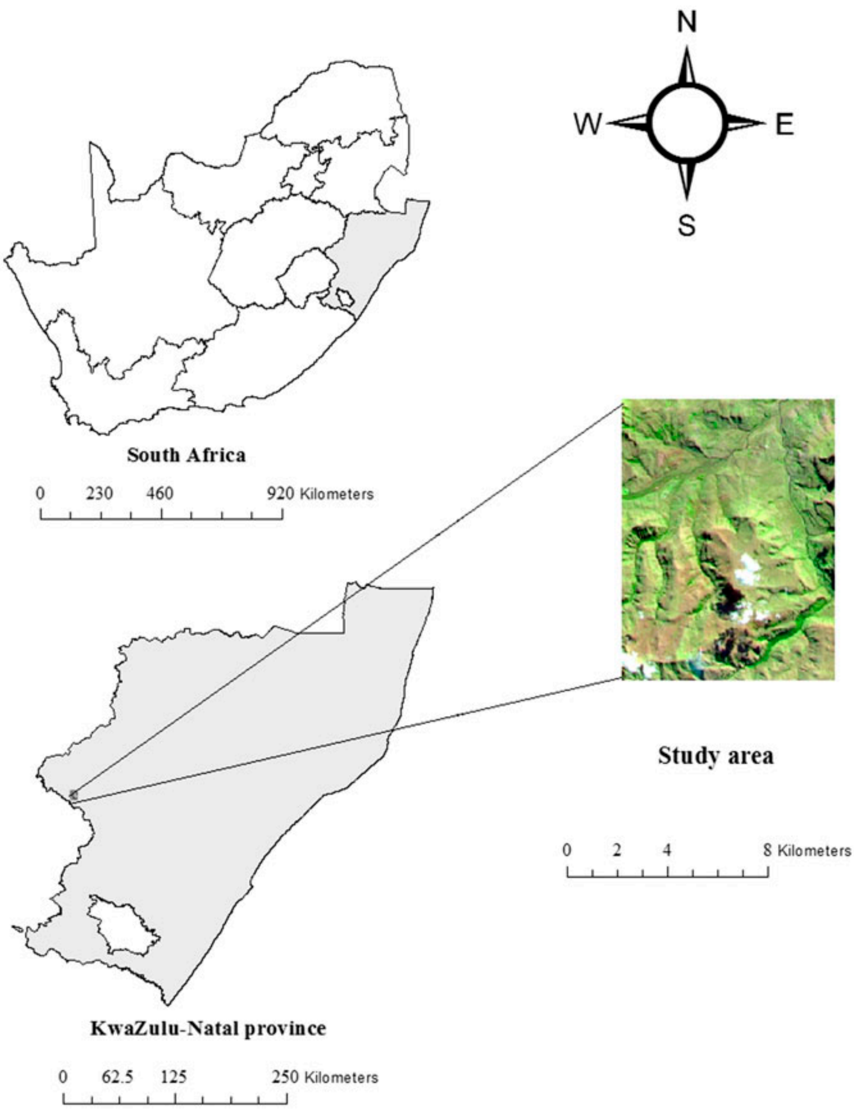

Figure 1. The location of Cathedral Peak in KwaZulu-Natal Province of South Africa.

Table 1. SPOT 5 scene parameters used in mapping the spatial distribution of decreaser and increaser species in the study area.

Spectral band	Pixel size (m)	Band width (nm)
B1: green	10	500–590
B2: red	10	610–680
B3: NIR	10	780–890
B4: MIR	20	1580–1750

Note: NIR = Near infrared, MIR = Middle infrared.

(Figure 1). The altitude varies from about 1860 to 2070 m above sea level (Everson 2001). Climate is typically wet and humid in summer (maximum temperature ranges from 18 to 26 °C) and dry cold in winters (minimum temperature ranges from 3 to 14 °C). Precipitation falls during the October to March summer months, with a mean annual value of 1350 mm (Nel & Sumner 2005). Soil consists of basalt-derived silt clays in the low areas, and shales, sandstone and mudstone on slopes and plateaux (Everson 2001; Govender & Everson 2005). Soils are classified as lateritic red and yellow earths, grading into heavy black soils (Granger & Schulze 1977). These soils are often highly leached, acidic and structure-less, with a mean depth of 0.8 m (Schulze 1975; Everson 2001). The study area is extensively covered by grassland and falls into the Montane Belt. The vegetation falls under the Moist Highland Sourveld, KwaZulu-Natal Bioresource Group 8 (Camp 1997), and is characterized by high variability in accordance to differences in altitude, temperature, soil and rainfall. The common species are: *Themeda triandra*, *Hyparrhenia hirta*, *Monocymbium ceresiiforme*, *Sporobolus pyramidalis*, *Eragrostis plana* and *Digitaria tricholaenoides* (Mucina & Rutherford 2006). In this study, to monitor grassland condition, two dominant species – *T. triandra and H. hirta* – were selected based on their relative abundance and distribution.

2.2. Image acquisition and preprocessing

A SPOT 5 satellite image (path 111) covering the KwaZulu-Natal wildlife conservation area of Cathedral Peak was acquired on 4 December 2010 (4:50 GTM). The image has four wavelengths in the green, red, near-infrared and med-infrared portions of the electromagnetic spectrum (EMS) and a 10 m spatial resolution (Table 1). The image was provided as SPOT 5 level 1A which was calibrated to sensor radiance. The scene radiance was therefore atmospherically corrected and transformed to at canopy reflectance using ATCOR3-r (Atmospheric/Topographic Correction-rugged terrain) algorithm built in ERDAS software. Atmospheric correction removes the effects of atmospheric conditions such as haze, mist and fog across the scene (Lillesand & Kiefer 2001). ATCOR3-r is based on radiative-transfer codes (MODTRAN-4) which computes the observed spectral radiance from the sensor for specified atmospheric features, sun angle and surface reflectance function (Berk et al. 1998). A SPOT 5 image was then geometrically registered to the Universal Transverse Mercator coordinates system using ground control points (GCPs). We further subset the image to cover an extent of about 11,000 × 8500 km (Figure 1).

2.3. Field data collection

Within seven days of image acquisition, GCPs, soil samples and the spatial distribution of the decreaser dominated by *T. triandra*, and the increaser dominated by the *H. hirta* (Figure 2) were collected from the conserved area of Cathedral Peak. 30 × 30 m plots were defined to represent respective dominant (over 80%) decreaser and increaser grass species. The point of each plot was recorded to obtain accurate reference data (GCP). In total, 30 sample plots were recorded for each target grass community ($n = 2$). Thirty GCPs from other classes in the study area viz., bare soil, settlements, and water bodies were also collected. The plots and other class GCPs were then used to generate regions of interest. Soil samples were collected at 0–15 cm depth using a soil-sampling auger from the 60 plots. In each plot, three samples were collected and thoroughly mixed to form one composite sample for analysis. A field label which indicated soil sample plot

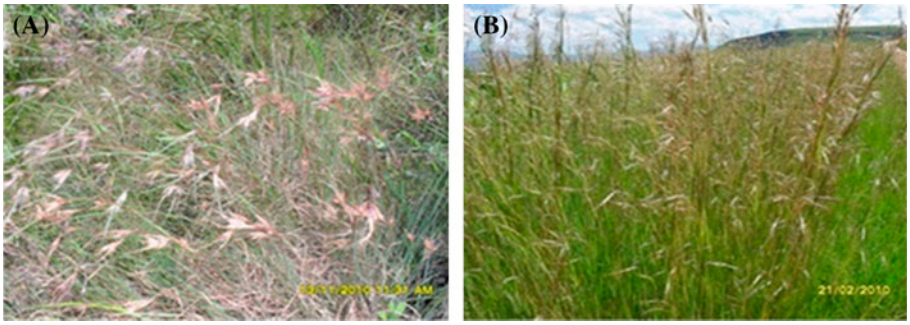

Figure 2. The most common grass species in the study area used for mapping grassland degradation: (A) *Themeda triandra* (B) *Hyparrhenia hirta.*

number, species name and GPS coordinates was attached to each sample bag. All labelled bags were stored in dry conditions until they were transported to the laboratory for analysis.

2.4. Soil physical and chemical analysis

The soil samples were analysed for available P, soluble K, Ca, total N, micronutrients (Mg, Zn, Mn and Cu), soil pH and organic carbon (OC). Available P was estimated by extraction using ammonium bicarbonate solution (0.25 buffered at pH 8.0), and calorimetrically determined (The Non-Affiliated Soil Analysis Work Committee 1990). Total K, Ca, Mg and micronutrients (Zn, Mn and Cu) were determined using electrothermal flame atomic absorption spectrometry (The Non-Affiliated Soil Analysis Work Committee 1990). The soil pH was measured in soil suspension (1 soil: 5_{KCl}) using a digital pH-metre with a connected glass electrode probe. Total N and OC were determined by mid-infrared spectroscopy (McCarty et al. 2010).

2.5. Statistical analysis

2.5.1. A paired sample t*-test*

A *t*-test was performed with 99% confidence levels ($p \leq 0.01$) to test the research hypothesis that there were significant differences in soil properties ($n = 10$) between decreaser and increaser species sites. We tested the research null hypothesis H_0: $\mu1 = \mu2 = \mu3 = \mu n$ vs. the alternative hypothesis H_a: $\mu1$ # $\mu2$ # $\mu3$ # μn, where $\mu1$, $\mu2$, $\mu3$, μn are the mean values for a specific element (i.e. P, K, Ca, Mg, pH, Zn, Mn, Cu, OC and N).

2.5.2. Random forest (RF) classifier

Random forest (RF) classifier is a machine learning algorithm developed by Breiman (2001). The algorithm generates multiple bootstrap samples (approximately 66% each) from the original training data-set with a replacement to create multiple classification trees (*ntree*) and uses the entire forest as a complex composite classifier. About 33% of the original data are left out from each bootstrapped sample, known as out-of-bag (OOB) data (Breiman 2001; Liaw & Wiener 2002; Archer & Kimes 2008). Each tree in the RF is independently grown to maximum size without any pruning. In each tree, RF uses randomness in the classification process by selecting a random subset of candidate features (*mtry*) to determine the best split at each node in a tree (Breiman 2001). The classification trees in the ensemble then vote by plurality on the correct classification. The RF algorithm provides three independent variable importance measures: the permutation accuracy importance measure, the Gini importance measure and the number of times that each variable is selected (Breiman 2001). The permutation accuracy importance measure is considered to be the best by-product in RF because of its ability to assess the variable importance which relies on mean decreases in accuracy as measured using OOB samples (Breiman 2001). The OOB error produces a measure of the importance of the variables by comparing how much the OOB error of estimate increases when a variable is permutated whilst all other variables are left unchanged (Archer & Kimes 2008). In the present study, the importance of each SPOT 5 band and the edaphic factors in mapping grassland degradation was determined using OOB estimates of classification error. The difference between the misclassification rate for the predicted and original OOB sample observations over all trees was then averaged to measure the importance of the variables (SPOT 5 bands, combined SPOT 5 bands and edaphic factors). The average was then used as a ranking index (mean decrease in accuracy) to identify the most suitable variables for mapping grassland degradation. A more detailed statistical description of RF algorithm can be found among others Breiman (2001) and Touw et al. (2012).

In the present study, the R software library (Liaw & Wiener 2002) developed in the R package for statistical analysis was used to carry out the classification algorithm (R Development Core Team 2008). The GCPs were used to create class signatures using SPOT 5 data. Furthermore, data of significant edaphic factors corresponding to the two grass species were employed as ancillary variables to test their usefulness in enhancing the accuracy of mapping the decreaser and increaser grass species using SPOT 5 image. In order to obtain high classification accuracy, the *mtry* and *ntree* parameters were optimized based on the OOB estimate of error (Breiman 2001). The *ntree* values were tested from a default setting of 500–10,000 trees with intervals of 500, while the *mtry* values were optimized by creating RF ensembles using all possible *mtry* values ($n = 4$) and ($n = 7$) for SPOT 5 bands and combined SPOT 5 bands and edaphic factors, respectively.

2.6. Classification accuracy assessment

A number of researchers (e.g. Lawrence et al. 2006; Prasad et al. 2006; Hsueh et al. 2013) have noted that the RF internal OOB error rate is a reliable and unbiased measure of misclassification because the RF algorithm generates OOB test data independently from the training data. On the other hand, the OOB error estimate is useful when a small number of field observations are used for analysis. Since we only collected 30 GCPs from each grass species site, bare soil, settlements and water bodies ($n = 150$ in total), the OOB error procedure was implemented to assess the performance of the RF classifier in mapping grassland condition. A confusion matrix was then constructed to compare the true class with the class assigned by the RF classifier and to calculate the overall accuracy as well as the producer and user's accuracies based on the OOB data. Furthermore, to assess the reliability of classification results, the quantity and allocation disagreements (QD and AD, respectively) by Pontius and Millones (2011) were calculated from the cross-tabulation matrices. The quantity disagreement is the absolute dissimilarity between the number of test (reference) observations and the predicted ones, while the allocation disagreement determines the number of predicted classes that deviate spatially from the reference samples.

3. Results

3.1. Variable selection using t-test

The *t*-test results indicated that there were significant differences ($p \leq 0.01$) between increaser and decreaser sites in some edaphic variables; P ($t = 5.26$), OC ($t = 6.35$), and N ($t = 8.56$). Decreaser sites showed significantly lower mean values than increaser sites for the three edaphic factors (Table 2). The table also shows that the variability in these significant edaphic variables was higher in decreaser than increaser sites. However, other elements (K, Ca, Mg, Zn, Mn, Cu and pH) showed no significant differences ($p \geq 0.01$) between the two sites. The significant soil variables ($n = 3$) were then used for subsequent analysis.

3.2. Optimizing RF parameters (ntree and mtry)

The results of optimizing RF parameters indicated optimal *mtry* and *ntree* values of 2 and 1000 for SPOT 5 bands, and 3 and 6500 for combined SPOT 5 and three edaphic variables, respectively (Figure 3). This is because the OOB error rates of 25 and 11.67% were relatively lower for SPOT 5 bands and combined SPOT 5 bands and three edaphic variables (Figure 3). These optimal values of *ntree* and *mtry* were used for all subsequent analyses.

3.3. Variable importance

SPOT 5 bands and combined SPOT 5 bands and three edaphic variables were ranked according to their importance in mapping grassland degradation. A mean decrease in accuracy which was based on the OOB error rate was used as a criterion for the ranking. Figure 4 shows that the most useful bands were B3 (NIR) and B1 (green), while the most useful edaphic factors were N and OC.

3.4. Classification accuracy assessment

An overall accuracy of 75.30% (QD = 2 and AD = 23) was obtained when using SPOT 5 bands (Table 3). The results indicated that there was considerable improvement in accuracy of mapping grassland degradation based on the spatial distribution of indicator species when the significant three edaphic factors were combined with SPOT 5 bands.

Table 2. Descriptive statistics of edaphic factors (Phosphorus: P, Organic carbon: OC, and Nitrogen: N) that showed significant differences ($p \leq 0.01$) between the sites ($n = 30$) of decreaser and increaser grass species.

	Decreaser grass species		Increaser grass species	
Edaphic factor	Mean	SD	Mean	SD
P (mg P kg^{-1})	06.08	03.49	12.97	02.46
OC (%)	03.27	02.42	07.65	1.47
N (%)	00.28	0.27	00.77	0.15

Note: SD = Standard deviation.

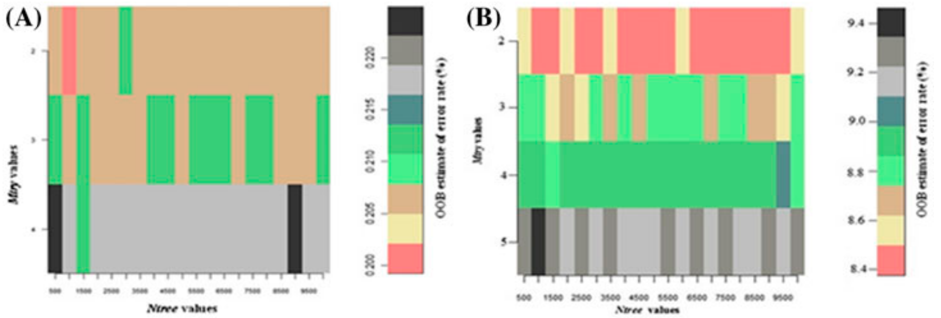

Figure 3. Optimizing RF parameters (*ntree* and *mtry*) using the OOB error rate. The *ntree* (1000 and 6500) and default setting of *mtry* (2 and 3) yielded the lowest OOB error rate (25 and 11.67%): SPOT 5 bands (A) and combined SPOT 5 bands and three edaphic variables (B).

The improvements included both classification accuracy (88.60%) and disagreement parameters (QD = 1 and AD = 11) (Table 4). In classifying individual classes, the results of the confusion matrices show that the classes that produced the lowest error rate were DE (3.30% when SPOT 5 bands were used and 10.3% when combined SPOT 5 bands and edaphic factors were used) followed by WB (6.90% when SPOT 5 bands were used and 25.00% when combined SPOT 5 bands and edaphic factors were used). On the other hand, IN and SM classes produced the highest error rate (16.70% when SPOT 5 bands were used and 31.00% when combined SPOT 5 bands and edaphic factors were used). Furthermore, results showed that very few decreaser and increaser sample points were confused with each other. Figure 5 shows a classification map of the study area. It can be asserted that the area is undergoing degradation as indicated by the domination of increaser grass species, except a few patches towards the middle and south west of the study area.

4. Discussion

This paper aimed at mapping grassland degradation in Cathedral Peak, South Africa using indicators of grass species. The indicator species were based on spatial distribution of a decreaser community, dominated by *T. triandra*, and increaser community

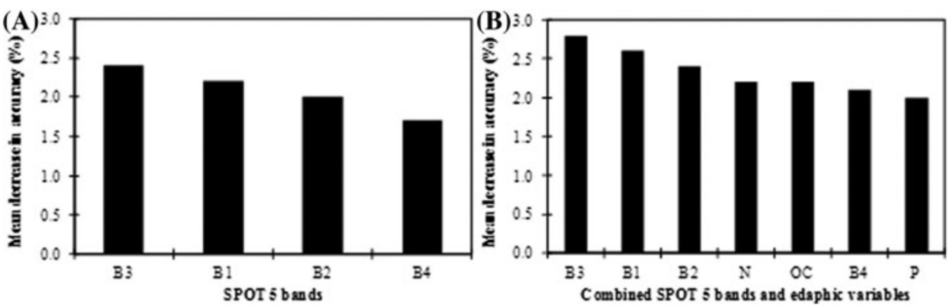

Figure 4. The ranking of the variables used in the classification according to their importance in mapping decreaser and increaser grass species as well as other classes in the study area using the random forest algorithm. (A) SPOT 5 bands and (B) SPOT 5 bands and three edaphic factors.

Table 3. Confusion matrix of random forest classification algorithm using SPOT 5 four bands for class sample data: decreaser (DE), increaser (IN), bare soil (BS), settlement (SM) and water body (WB). The confusion matrix includes overall accuracy (AC), quantity disagreement (QD), allocation disagreement, commission error (CE), omission error (OE), user's accuracy (UA) and producer's accuracy (PA).

Class	DE	IN	BS	SM	WB	Row total	CE	UA (%)
DE	26	1	0	1	1	29	10.3	89.7
IN	2	20	3	3	2	29	31.0	69.0
BS	0	4	22	3	1	31	29.0	71.0
SM	0	3	3	21	2	29	27.6	72.4
WB	2	2	2	2	24	32	25.0	75.0
Column total	30	30	30	30	30	150		
OE	13.3	35.5	26.7	25	22.6			
PA (%)	86.7	64.5	73.3	75	77.4			
OA (%) = 75.30								
QD = 2								
AD = 23								

Table 4. Confusion matrix of random forest classification algorithm using combined SPOT 5 bands and three edaphic variables (N, P and OC) for class sample data: decreaser (DE), increaser (IN), bare soil (BS), settlement (SM) and water body (WB). The confusion matrix includes overall accuracy (AC), quantity disagreement (QD), allocation disagreement (AD), commission (CE), omission error (OE), user's accuracy (UA) and producer's accuracy (PA).

Class	DE	IN	BS	SM	WB	Row total	CE	UA (%)
DE	29	1	0	0	0	30	03.3	96.7
IN	1	25	1	1	1	30	16.7	83.3
BS	0	1	27	2	1	31	12.9	87.1
SM	0	2	2	25	1	30	16.7	83.3
WB	0	1	0	2	27	29	06.9	93.1
Column total	30	30	30	30	30	150		
OE	3.3	19.4	12.9	13.8	6.9			
PA (%)	96.7	80.6	87.1	86.2	93.1			
OA (%) = 88.60								
QD = 1								
AD = 11								

dominated by the *H. hirta*s. To achieve this goal, an investigation of the utility of SPOT 5 satellite imagery and RF classification algorithm was undertaken. Edaphic variables were used to improve the accuracy of mapping grassland degradation.

A *t*-test results show that there were significant differences ($p \leq 0.01$) in some soil variables (P, N and OC) between decreaser and increaser grass sites. The results obtained in this study indicate that RF classification algorithm performed relatively better in analysing SPOT 5 data for monitoring grassland degradation. The results of the RF parameters (*mtry* and *ntree*) showed that a large number of *ntree* (1000 and 2000), and the default setting of *mtry* (2 and 3) are the best choices that yielded the lowest OOB error rates (25 and 11.67%) for bands and combined bands and significant edaphic variables, respectively. These results are consistent with among others Adam et al. (2012) which mentioned the importance of using a large number of *ntree* and default *mtry* value for discriminating vegetation species.

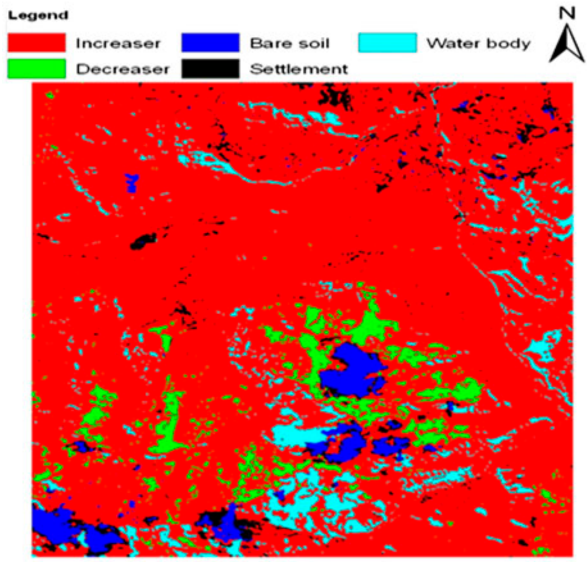

Figure 5. A classification map obtained using the random forest algorithm and SPOT 5 four bands only.

Mapping the spatial distribution of decreaser and increaser grasses and other classes in the study area demonstrated the utility of SPOT 5 data in monitoring grassland degradation, with an overall accuracy of 75.30%. The importance of SPOT 5 bands in mapping grassland degradation using RF classifier is presented in Figure 4. Specifically, the information of grassland degradation was contained in the NIR (780–890 nm), green (500–590 nm), red (610–680 nm) and MIR (1580–1750 nm) regions of the EMS. Mansour and Mutanga (2012) and Omar (2010) highlighted the value of NIR and visible regions in discriminating vegetation species. Since NIR and green portions of the EMS are water and chlorophyll absorption features (Kumar et al. 2001), the two species (decreaser and increaser) could vary in their leaf water and chlorophyll (greenness, Figure 2) contents. Different plant canopy geometries can also result in different spectral properties (Lillesand & Kiefer 2001).

The grassland condition was further mapped with a relatively higher overall accuracy (88.60%) when three edaphic factors (N, P and OC) were combined with SPOT 5 data. Since the three soil variables differed significantly within increaser and decreaser sites, in terms of the improvement in classification accuracy due to their inclusion was therefore expected. This finding was reinforced by the result in Table 2 which shows that lower fertility sites were associated with the decreaser species and the higher fertility sites with the increaser species. The finding is consistent with van Oudtshoorn (1999) and Dobarro et al. (2010) who reported that the increaser grass species are an indicator to land degradation (low land productivity). In addition, these three soil variables are essential nutrients for plant growth and development (Fageria 2009). Hence, their variation could have resulted in distinct plant physiological and morphological features that manifested in different spectral features detected by SPOT 5 sensor. This result supports previous recommendations suggesting that improved classification accuracy is obtainable by incorporating vegetation-related environmental variables with multispectral data (DomaÇ & Süzen 2006). Furthermore, these results are in line with our initial hypothesis that soil edaphic factors combined with multispectral data would improve mapping of decreaser and increaser grass species.

Our study mapped two grass species; *H. hirta and T. triandra*. The former is associated with disturbed landscapes (such as old-land), has a poor grazing value and has a bitter taste. This species is abundant at low to mid-altitudes along rivers, roads and around settlements, particularly in the poor soil of upland regions of South Africa (Cilliers & Bredenkamp 2000). On the other hand, the later is a widely distributed in undisturbed areas at high attitude (1300–3000 m above sea level). Due to its high nutritive and animals palpability, it is considered one of the most valuable range species. Therefore, the presence of the *T. triandra* species is a good indicator of a well-managed healthy rangeland (van Oudtshoorn 1999). It is known that soil spectral properties are difficult to depict using satellite data during grass growing season; the relative abundance of decreaser and increaser grass species could therefore be used as indicator of soil fertility. In a recent study in South Africa, Dlamini et al. (2014) found that land degradation, defined as a reduction in aerial grass cover, impacted significantly on soil N and OC stocks of grassland. Future studies should model areas that might be infested by *H. hirta* using parameters associated with this species and geospatial modellers. This would enable rangeland managers to make an early decision or policy for maintaining healthy grassland conditions before increaser invasion takes place in the non-degraded sites. It is found that the study area was undergoing degradation (Figure 5), future studies should therefore look at other causes, rather than edaphic variables, that might have led to such degradation. Since we utilized freely available remotely sensed data from the commonly used SPOT 5 sensor, models developed in the present study could further be tested for an operational application of monitoring grasslands degradation in areas with similar conditions in South Africa.

5. Conclusions

This study has shown that soil P, OC and N contents vary significantly between sites with decreaser (*T. triandra*) and increaser (*H. hirta*) grass species. The spatial distribution of these grass species has a high potential for classifying degraded and non-degraded grassland sites. It also found that integrating edaphic variables (P, OC and N) with SPOT 5 multispectral data resulted in an improved classification accuracy compared to multispectral data only. SPOT 5 data are potentially cost-effective and applicable in operational decreaser and increaser species classification.

Overall, our study suggests that multispectral remote sensing and edaphic factors can effectively provide a baseline for the spatial distribution of decreaser and increaser species for monitoring and assessing grassland degradation. Furthermore, detailed information on the spatial distribution of decreaser and increaser species can help environmental policy-makers to formulate conservation strategies for sustainable rangeland management.

Acknowledgements

Thanks are due to the South African National Space Agency (SANSA) for making high-quality SPOT 5 data available free of charge for this study. The authors thank Dr Irene Bame and Khatab Abdallah for their advices and assistance on matters pertaining to soil analysis.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by UKZN and the National Research Foundation (NRF) research grant.

References

Adam EM, Mutanga O, Rugege D, Ismail R. 2012. Discriminating the papyrus vegetation (*Cyperus papyrus* L.) and its co-existent species using random forest and hyperspectral data resampled to HYMAP. Int J Remote Sens. 33:552–569.

Archer K, Kimes R. 2008. Empirical characterization of random forest variable importance measures. Comput Stat Data Anal. 52:2249–2260.

Berk A, Bernstein LS, Anderson GP, Acharya PK, Robertson DC, Chetwynd JH, Adler-Golden SM. 1998. MODTRAN cloud and multiple scattering upgrades with application to AVIRIS. Remote Sens Environ. 65:367–375.

Breiman L. 2001. Random forests. Mach Learn. 45:5–32.

Camp K. 1997. The bioresource groups of KwaZulu-Natal, CEDARA Report N/A/97/6. Pietermaritzburg: Department of Agriculture and Environmental Affairs.

Chapin FS. 1980. The mineral nutrition of wild plants. Annu Rev Ecol Syst. 11:233–260.

Chastain RA Jr., Struckhoff MA, He HS, Larsen DR. 2008. Mapping vegetation communities using statistical data fusion in the Ozark national scenic river ways, Missouri, USA. Photogramm Eng Remote Sens. 74:24–264.

Cho MA, Mathieu R, Asner GP, Naidoo L, van Aardt J, Ramoelo A, Debba P, Wessels K, Main R, Smit IPJ, Erasmus B. 2012. Mapping tree species composition in South African savannas using an integrated airborne spectral and LiDAR system. Remote Sens Environ. 125: 214–226.

Cilliers SS, Bredenkamp GJ. 2000. Vegetation of road verges on an urbanisation gradient in Potchefstroom, South Africa. Landscape Urban Plan. 46:217–239.

Dlamini P, Chivenge P, Manson A, Chaplot V. 2014. Land degradation impact on soil organic carbon and nitrogen stocks of sub-tropical humid grasslands in South Africa. Geoderma. 235–236:372–381.

Dobarro I, Valladares F, Peco B. 2010. Light quality and not quantity segregates germination of grazing increasers from decreasers in Mediterranean grasslands. Acta Oecol. 36:74–79.

DomaÇ A, Süzen ML. 2006. Integration of environmental variables with satellite images in regional scale vegetation classification. Int J Remote Sens. 27:1329–1350.

Everson C. 2001. The water balance of a first order catchment in the montane grasslands of South Africa. J Hydrol. 241:110–123.

Fageria NK. 2009. The use of nutrients in crop plants. New York (NY): Taylor and Francis Group.

Govender M, Everson CS. 2005. Modelling streamflow from two small South African experimental catchments using the SWAT model. Hydrol Processes. 19:683–692.

Granger J, Schulze R. 1977. Incoming solar radiation patterns and vegetation response: examples from the natal drakensberg. Plant Ecol. 35:47–54.

Harvey K, Hill G. 2001. Vegetation mapping of a tropical freshwater swamp in the Northern Territory, Australia: a comparison of aerial photography, Landsat TM and SPOT satellite imagery. Int J Remote Sens. 22:2911–2925.

Hoffman M, Todd S. 2000. A national review of land degradation in South Africa: the influence of biophysical and socio-economic factors. J S Afr Stud. 26:743–758.

Hsueh H-M, Zhou D-W, Tsai C-A. 2013. Random forests-based differential analysis of gene sets for gene expression data. Gene. 518:179–186.

Kumar L, Schmidt KS, Dury S, Skidmore AK. 2001. Imaging spectrometry: basic principles and prospective applications. Dordrecht: Kluwer Academic Press. Chapter 5, Imaging spectrometry and vegetation science; p. 111–155.

Lawrence RL, Wood SD, Sheley RL. 2006. Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (randomForest). Remote Sens Environ. 100:356–362.

Lees BG, Ritman K. 1991. Decision-tree and rule-induction approach to integration of remotely sensed and GIS data in mapping vegetation in disturbed or hilly environments. Environ Manage. 15:823–831.

Liaw A, Wiener M. 2002. Classification and regression by randomForest. R News. 2:18–22.

Liberti M, Simoniello T, Carone MT, Coppola R, D'Emilio M, Macchiato M. 2009. Mapping badland areas using LANDSAT TM/ETM satellite imagery and morphological data. Geomorphology. 106:333–343.

Lillesand TM, Kiefer RW. 2001. Remote sensing and image interpretation. 4th ed. New York (NY): Wiley.

Mansour K, Mutanga O. 2012. Classifying increaser species as an indicator of different levels of rangeland degradation using WorldView-2 imagery. J Appl Remote Sens. 6:1–18.

Mansour K, Mutanga O, Everson T, Adam E. 2012. Discriminating indicator grass species for rangeland degradation assessment using hyperspectral data resampled to AISA Eagle resolution. ISPRS J Photogramm Remote Sens. 70:56–65.

Marschner H. 1993. Mineral nutrition of higher plants. London: Academic Press.

Marschner H, Rimmington G. 1988. Mineral nutrition of higher plants. Plant Cell Environ. 11:147–148.

Martínez-López J, Carreño MF, Palazón-Ferrando JA, Martínez-Fernández J, Esteve MA. 2014. Remote sensing of plant communities as a tool for assessing the condition of semiarid Mediterranean saline wetlands in agricultural catchments. Int J Appl Earth Obs Geoinformation. 26:193–204.

McCarty GW, Reeves J, Yost R, Doraiswamy PC, Doumbia M. 2010. Evaluation of methods for measuring soil organic carbon in West African soils. Afr J Agric Res. 5:2169–2177.

Muchoney D, Haack B. 1994. Change detection for monitoring forest defoliation. Photogramm Eng Remote Sens. 60:1243–1251.

Mucina L, Rutherford MC. 2006. The vegetation of South Africa, Lesotho and Swaziland. Strelitzia. 19:616–657.

Nel W, Sumner PD. 2005. First rainfall data from the KZN Drakensberg escarpment edge (2002 and 2003). Water S Afr. 31:399–402.

Omar H. 2010. Commercial timber tree species identification using multispectral WorldView2 data. Longmont (CO): Digitalglobe® 8 Bands Research Challenge; p. 2–13.

Pontius RG, Millones M. 2011. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. Int J Remote Sens. 32:4407–4429.

Prasad AM, Iverson LR, Liaw A. 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems. 9:181–199.

Rignot E, Salas WA, Skole DL. 1997. Mapping deforestation and secondary growth in Rondonia, Brazil, using imaging radar and thematic mapper data. Remote Sens Environ. 59:167–179.

SANSA. 2012. SPOT 5 South African mosaic [Internet]. South African National Space Agency; [cited 2014 Aug 15]. Available from: http://www.sansa.org.za/earthobservation/services

Schmidtlein S, Sassin J. 2004. Mapping of continuous floristic gradients in grasslands using hyperspectral imagery. Remote Sens Environ. 92:126–138.

Schulze RE. 1975. Catchment evapotranspiration in the Natal Drakensberg [dissertation]. Pietermaritzburg: University of Natal.

The Non-Affiliated Soil Analysis Work Committee. 1990. Handbook of standard soil testing methods for advisory purposes. Pretoria: Soil Science Society of South Africa.

Tainton NM. 1999. Veld management in South Africa. Pietermaritzburg: University of Natal Press. Chapter 3, The grassland biome; p. 25–33.

Thenkabail P, Enclona E, Ashton M, Van Der Meer B. 2004. Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications. Remote Sens Environ. 91:354–376.

Thompson M. 1996. Standard land-cover classification scheme for remote-sensing applications in South Africa. S Afr J Sci. 92:34–42.

Touw WG, Bayjanov JR, Overmars L, Backus L, Boekhorst J, Wels M, Hijum SAFTV. 2012. Data mining in the life sciences with random forest: a walk in the park or lost in the jungle? Brief Bioinform. 14:315–326. doi:10.1093/bib/bbs034.

Underwood E, Ustin S, DiPietro D. 2003. Mapping nonnative plants using hyperspectral imagery. Remote Sens Environ. 86:150–161.

Ustin SL, Jacquemoud S, Palacios-Orueta A, Li L, Whiting ML. 2009. Recent advances in remote sensing and geoinformation processing for land degradation assessment. London: CRC Press. Chapter 2, Remote sensing based assessment of biophysical indicators for land degradation and desertification; p. 15–44.
van den Berg L, Zeng Y. 2006. Response of South African indigenous grass species to drought stress induced by polyethylene glycol (PEG) 6000. S Afr J Bot. 72:284–286.
van Oudtshoorn FP. 1999. Guide to grasses of Southern Africa. Pretoria: Briza Publications.
Wessels K, Prince S, Frost P, van Zyl D. 2004. Assessing the effects of human-induced land degradation in the former homelands of northern South Africa with a 1 km AVHRR NDVI time-series. Remote Sens Environ. 91:47–67.