Monitoring grass nutrients and biomass as indicators of rangeland quality and quantity using random forest modelling and WorldView-2 data

Abel Ramoelo[1,3*]**, M.A. Cho**[1]**, R. Mathieu**[1]**,S. Madonsela**[1]**, R. van de Kerchove**[2]**, Z. Kaszta**[2]**, E. Wolff**[2]

[1]Natural Resource and Environment, Council for Scientific and Industrial Research (CSIR), Pretoria, South Africa
[2]IGEAT, Université Libre de Bruxelles (ULB), Belgium
[3]Risk and Vulnerability Assessment Centre, University of Limpopo, Sovenga, South Africa
*Correspondinig author email: aramoelo@csir.co.za or abel.ramoelo@gmail.com

Abstract

Land use and climate change could have huge impacts on food security and the health of various ecosystems. Leaf nitrogen (N) and above-ground biomass are some of the key factors limiting agricultural production and ecosystem functioning. Leaf N and biomass can be used as indicators of rangeland quality and quantity. Conventional methods for assessing these vegetation parameters at landscape scale level is time consuming and tedious. Remote sensing provides a bird-eye view of the landscape, which creates an opportunity to assess these vegetation parameters over wider rangeland areas. Estimation of leaf N has been successful during peak productivity or high biomass and limited studies estimated leaf N in dry season. The estimation of above-ground biomass has been hindered by the signal saturation problems using conventional vegetation indices. The objective of this study is to monitor leaf N and above-ground biomass as an indicator of rangeland quality and quantity using WorldView-2 satellite images and random forest technique in the north-eastern part of South Africa. Series of field work to collect samples for leaf N and biomass were undertaken in March 2013, April or May 2012 (end of wet season) and July 2012 (dry season). Several conventional and red edge based vegetation indices were computed. Overall results indicate that random forest and vegetation indices explained over 89% of leaf N concentrations for grass and trees, and less than 89% for all the years of assessment. The red edge based vegetation indices were among the important variables for predicting leaf N. For the biomass, random forest model explained over 84% of biomass variation in all years, and visible bands including red edge based vegetation indices were found to be important. The study demonstrated that leaf N could be monitored using high spatial resolution with the red edge band capability, and is important for rangeland assessment and monitoring.

Keywords: rangeland quality, leaf nitrogen, biomass, random forest model, WorldView-2, red edge band

Introduction

Historically, rural African people were nomadic in nature and rarely affected by stringent spatial planning and zoning for natural resource use (e.g. grazing camps). Now, rural people are no longer traveling long distances for grazing and for agricultural production. This current situation warrant for a proper management of land because land degradation is continuing to be a problem for rural systems. Land degradation caused by changes in anthropogenic and climatic factors pose threat to the rural economy. Rural people rely on land for grazing for livestock production, fuelwood for heating and cooking as well as crop production which are central to the rural economy and livelihood (Shackleton et al., 2002). It is crucial to develop information appropriate to safeguard natural resources while improving livestock production, hence the rural economy.

Leaf nitrogen (N) as a measure of rangeland quality is one of the key factors limiting agricultural production and ecosystem functioning or health. Leaf N is a good indicator of photosynthetic capacity of plants. For example, leaf N is a constituent of the Nicotinamide Adenine Dinucleotide Phosphate (NADPH) co-enzyme used in the Photosystem I process. (Devlin, 1975). In the light-independent Calvyn-Benson Cycle, both NADP and adenosine triphosphate (ATP) are used with the enzyme ribulose biphosphate carboxylase (RuBisCo) to bind carbon dioxide and form sugars, sucrose, starches, proteins and structural components in cell tissues of the plants (Devlin, 1975). Grass quantity is measured by above-ground biomass (mass per unit area). Grass N (both trees and grasses) and biomass can be used as an indicator of rangeland quality and quantity respectively, which could provide information for planning and management by farmers, decision makers, land planners and managers.

Leaf N is related to protein (Clifton et al., 1994; Devlin, 1975) and biomass play a crucial role in understanding the feeding patterns and distribution of wildlife and livestock (Drent and Prins, 1987; McNaughton, 1988, 1990). Several studies postulated that large herbivores (e.g. buffalo, elephant, cattle) also feed and concentrate in areas of high nutritious grasses (Grant and Scholes, 2006; Owen-Smith and Danckwerts, 1997), and require high biomass. Small herbivore, on the other hand, such as impala, are selective grazers, feeding on nutritious grasses (Bothma, 1989) and are not bulk grazers. Nutritious grasses are mainly found underneath the trees, on termite mounts and sodic sites (Grant and Scholes, 2006; Ludwig et al., 2008; Ludwig et al., 2001; Treydte et al., 2007). The major drivers of grass nutrient concentrations and biomass are geology and topography. Basalt and gabbro geological types are known to support high grass nutrients and biomass, because they are easily weathered and rich in clay (over 50%) (Venter et al., 2003). Granite type of geology supports less nutritious grass, because they are not easily weathered and contain less clay. In addition to geology, topography plays a crucial role in the distribution of biomass and nutrients, Bottomlands or valleys have high nutritious grasses, while crest has less nutritious grasses and biomass. Understanding the feeding patterns and movement of wildlife and livestock does not only provide information for decision makers only but can also be used to model risk of disease transfusion (e.g. tuberculosis, foot and mouth) along the fence-line contrast between communal and protected areas (e.g. cattle and buffalo) (Bengis et al., 2003; Chaminuka et al., 2011).

Remote sensing data has an advantage of covering wider areas which provides an opportunity to assess grass N and biomass at landscape scale from protected to communal areas, as compared to conventional point-based assessments. Successful estimation of leaf N was achieved using field and airborne imaging spectroscopy which are mainly applicable at local scale (Huang et al., 2004; Mutanga and Skidmore, 2004a; Yoder and Pettigrew-Crosby, 1995). Conventional technique are widely used involving narrow vegetation indices

for the estimation of various biophysical (e.g. Leaf Area Index, biomass etc.) and biochemical (e.g. Leaf N, chlorophyll content etc.) parameters (Darvishzadeh et al., 2008; Dijkshoorn et al., 2008; Huang et al., 2004; Mutanga and Skidmore, 2004b; Ramoelo et al., 2013; Ren and Zhou, 2012; Yoder and Pettigrew-Crosby, 1995). The conventional indices (e.g. Normalized Difference Vegetation Index-NDVI) from broadband sensors such as Landsat cannot be used for estimating leaf N, but were successfully used for biomass. Landsat does not have a red edge band. Using conventional vegetation indices such as NDVI to predict biomass, there is a well-known problem of signal saturation during peak productivity (Mutanga and Skidmore, 2004b; Tucker, 1977). The amount of light that can be absorbed at the red region of the spectrum reaches a plateau during the peak productivity (Mutanga and Skidmore, 2004b; Tucker, 1977). Additionally, the NIR reflectance continue to increase, because addition of new leaves influence the multiple scattering (Kumar et al., 2001). The vegetation index (e.g. NDVI) changes slightly, while causing a poor relationship with biomass. This problem was solved by using narrow band vegetation indices (Mutanga and Skidmore, 2004b) and red edge inflation point (REIP) based on hyperspectral data.

The use of red-edge based vegetation indices provide better estimates of leaf biochemical (Cho et al., 2013; Ramoelo et al., 2012) and biomass (Mutanga et al., 2012). The lack of red-edge bands in satellite multispectral sensors mired the production of regional maps of leaf N and improved biomass maps. New space-borne sensors such as WorldView-2 and RapidEye designed with red-edge bands provide an opportunity for assessing leaf N at regional scale (over 5000 km^2). Ramoelo et al., (2012) and Cho et al., (2013) reported an earlier work of estimating leaf N using RapidEye. In the next few years, the European Space Agency (ESA) is planning to launch a new satellite (Sentinel-2), with spectral bands comparable to those of RapidEye and WorldView-2. Testing and development of this technology with high resolution sensor such as WorldView-2 will allow to better calibrate and validate nutrient and biomass models relying on coarser resolution sensors, available at more frequent time step and at subcontinent scale (e.g. Sentinel-2).

Using remote sensing, the extraction of leaf N and biomass has been limited to the use of statistical analysis. The most common technique are simple and stepwise multiple linear regression (Ramoelo et al., 2012). The latter suffers from overfitting and multi-collinearity (Curran, 1989; Kokaly et al., 2009). Machine learning techniques such as random forest (RF) (Mutanga et al., 2012) and artificial neural network (ANN) (Knox et al., 2011; Mutanga and Skidmore, 2004a; Skidmore et al., 2010) were found to be robust, and circumvents overfitting and multi-collinearity problem. RF and ANN were in most cases used for classification than for regression analysis (Adam et al., 2012). In regression, RF was found to be more robust than other parametric regression techniques (Mutanga et al., 2012). Mutanga et al., (2012) demonstrated that using WorldView-2 images and RF, biomass can be estimated during peak productivity. A question to be investigated here is can new very high resolution multispectral imagery with red-edge bands such as WorldView-2 be used as to develop an approach to move towards regional assessment and monitoring of leaf N and biomass as indicators of rangeland quality and quantity?.

Data collection and sampling

WorldView-2 images with 2m x 2m spatial resolution for April, July 2012 and March 2013 were acquired. March-April period was collected end of wet season to autumn, while July was collected in dry season (winter). Studies such as Ramoelo et al., (2012) and Skidmore et al., (2010) showed that the best period to estimate grass nutrients is during peak productivity (end of wet season-March-April). The pre-processing was done, i.e. orthorectification and atmospheric correction using PCI Geomatica software. The rational polynomial coefficients (RPC) based orthorectification method with 0th, first and second

order corrections for sensor orientation were tested. The accuracy was evaluated using the leave-one-out- cross validation with 18 ground control points. The 0th order RPC was selected as it yielded lower root mean square error which is less than an image pixel size (<2m). The digital elevation model used was the 90m NASA Shuttle Radar Topographic Mission (SRTM) database v4.1 and provided good orthorectification results. The atmospheric correction was done using the Atmospheric Correction for flat surfaces (ATCOR2) because the study area is characterized by flat to gentle undulating slopes.

Sample points for collecting grass and tree leaf samples were purposively located using the acquired WorldView-2 image. The fieldwork was undertaken within two weeks of image acquisition. For grasses additional points were added in case the pre-selected points were burnt or overgrazed. For the dry season campaign, most of the tree leaf samples were collected in the riparian zones, because of leaf shading by most inland trees. In each sample point, a 6m x 6m plot of homogeneous grass cover was defined, and two–three subplots of 50cm x 50cm were randomly selected. In each subplot, grass samples were collected and data on dominant grass species was recorded. Within the vicinity of the plot, big trees of various species were identified and 5 leaves representative of the canopy were collected. The grass and tree leaf samples were dried (80^{0}C for 24 hours) and taken to the laboratory for chemical analysis to retrieve leaf N at Bemlab PTY (LTD), Strand, Western Cape, South Africa. Collected grass samples (above-ground) in each subplot were weighed before and after drying to determine aboveground biomass, which was converted to g/m^2. The dried above-ground biomass (g/m^2) was used for further analysis, and hereto referred to as biomass.

Data analysis

Spectral reflectance from WorldView-2 images for different dates were extracted corresponding to each sample GPS point with leaf N and biomass value. A 3 x 3 pixel window was used to extract the spectra corresponding to a plot size of 6m x 6m, for a 2m pixel size. The reflectance from 9 pixels was then averaged to derive a single representative reflectance. Conventional and red-edge based vegetation indices such as simple ratio (SR) (Jordan, 1969) and the normalized difference vegetation index (NDVI) (Rouse et al., 1974) were computed.   Several band combinations were used and based on the band combinations NDVI was named NDVI (855 and 605), NDVI1 (855 and 660nm), NDVI2 (950 and 605nm) and NDVI3 (950 and 660nm) and similarly for red-edge based vegetation indices, NDVI-RE1 (855 and 725nm), NDVI-RE2 (950-725nm), NDVI-RE3 (725 and 605nm) and NDVI-RE4 (725 and 660nm). The same order was done for SR and similar bands were used as for NDVI, where SR1 will correspond to NDVI1, for example and so on. Additional indices based on MERIS Terrestrial Chlorophyll Index (MTCI) (Dash and Curran, 2004) was modified and computed, MTCI1 (660, 725 and 855nm) and MTCI2 (660, 725 and 950nm). MTCI1 and 2 are both red-edge based, with different near infrared band used. In total 26 variables were used for regression, with 8 bands and 18 vegetation indices.

The vegetation indices and bands totalling 26, were used as an input into the nonparametric random forest (RF) method (Breiman, 2001) to predict leaf N and biomass. RF was chosen because leaf N and biomass data were not normally distributed. Random forest was implemented from the random forest package programmed in the R statistical environment (Liaw and Wiener, 2002). This technique was successfully used with remote sensing to predict wetland species biomass (Mutanga et al., 2012), plant water content (Ismail and Mutanga, 2010) and for species classification (Adam et al., 2012; Ham et al., 2005). RF is a machine learning method developed to improve the classification and regression trees method (CART) by using large set of decision trees. RF builds each tree by using a deterministic algorithm selecting a random set of variables and a random sample from the

calibration data sets. There are three main variables such as *ntree*; number of regression trees grown based on a bootstrap sample of observation (the default value is 500 trees), *mtry*; number of predictors tested at each node (default is the square-root of the total number of variables) and *nodesize*; minimal size of the terminal nodes of the trees (the default value is one) was used and it needs to be well-defined (Mutanga et al., 2012). The optimization of the number of variables required to predict leaf N and biomass was determined using a recursive feature selection based on leave-one-out cross validation (LOOCV) (Diaz-Uriarte and Alvarez de Andres, 2006) and the root mean square error (RMSE). The conventional application of random forest is outlined below based on Beiman, (2001) and Mutanga et al., (2012);

    (i)   *ntree* bootstrap sample X_i (i= bootstrap iteration) are randomly drawn with replacement from original dataset (calibration), each containing approximately one-third of the elements of the calibration data sets X. The variables not included in X_i are called out-of-bag data (OOB) for that bootstrap sample.

    (ii)   for each bootstrap sample an unpruned regression tree is grown with the modification that at each node, one-third of the predictor is randomly selected and the best split among those variables is chosen.

    (iii) at each bootstrap iteration, the response value for data not included in the bootstrap sample (OOB data) is predicted and averaged over all trees (*ntree*)

    (iv) the importance of each predictor is measured by calculating the percent increase in mean square error when the OOB data for each variable is permuted, while others are unchanged. These variable importance values are then used to rank the predictors in terms of their relationship to response variables. The higher the variable of importance score or value the higher the importance of the particular variable in the model.

The validation of the model was done using LOOCV because of the small size of the available sample. In cross validation, samples are estimated by the remaining samples. For example, if there are 20 samples, each sample will be predicted by 19 samples iteratively to determine the performance of the model. Merits of the cross-validation are the capability to detect outliers and provide unbiased assessment of the prediction error (Efron and Gong, 1983). The statistic measure of precision and accuracy such as the coefficient of determination (R^2) and root mean square error (RMSE) were determined.

To further understand the non- and spatial relationship between the predicted biomass and leaf N, the Spearman rank correlation (Hollander and Wolfe, 1973) and cross-variogram analysis, respectively (Bivand et al., 2008), implemented in R statistical programming language. For the cross-variagram, 1000 points were randomly generated and 157 used to extract collocated biomass and leaf N values from various maps. Several variogram fitting models were investigated and the better fit model was selected to undertake cross-variogram analysis. The details on the workflow for applying cross-variogram can be found in Bivand et al., (2008).

Results and discussion

Leaf N and biomass

Leaf N concentration mean and variability in grass were higher (mean=0.70, CV=25%) in March 2013 during the end of the wet season compared to the dry season in July (mean=0.49, CV=14%) (Table 1). Leaf N concentrations trend in grass is also depicted by higher maximum values for the wet season (March 2013, max=1.21) compared to the dry

season (July 2012, max =0.72). When combining grass and tree leaf N concentrations, the variability trend remain the same, with CV= 67% in March 2013 and April 2012, and 63% in July 2012. Generally, the grass variability of leaf N was low in both years, but higher when combined with that of trees. Green vegetation is likely to have higher nutrient concentrations as compared to dry ones because of active photosynthetic activities portraying the state or health of the vegetation. During dry periods the nutrients are generally trans-located from the leaves to the roots. The high maximum values of leaf N are from trees as shown in Table 1.

Biomass variability is higher (CV=52%) in March 2013, as compared to April 2012 and July 2012 (CV=39 and 49%, respectively). On average, the highest biomass values were achieved in March 2013, and the lowest in July 2012. Higher herbaceous biomass in savanna is attributed to the productive season characterized by high rainfall. During the dry season, factors such as overgrazing, especially in the communal areas, and fire lead to the dwindling of biomass. More details of the descriptive statistics of biomass can be seen in Table 1.

The interaction between leaf N and biomass in grass is minimal e.g. Ramoelo et al., (2012). The correlation between leaf N and biomass in different seasons was low and not significant ($p < 0.05$). These results support the simultaneous estimation of leaf N and biomass using remote sensing.

Table 1: Descriptive statistics of leaf N (%) and biomass (g/m^2) for various seasons or dates

Data sets	min	max	mean	median	STDEV	CV (%)	SE	No. of samples
Leaf N (Grass+Trees)								
April-2012	0.49	2.63	0.89	0.65	0.59	67	0.11	28*
Jul-2012	0.40	2.02	0.83	0.53	0.52	63	0.08	41*
March 2013	0.51	2.96	1.11	0.78	0.74	67	0.11	48*
Leaf N (Grass only)								
April-2012	0.49	0.78	0.61	0.60	0.09	15	0.02	21
Jul-2012	0.40	0.72	0.49	0.48	0.07	14	0.01	26*
March 2013	0.51	1.21	0.70	0.65	0.18	25	0.03	34*
Biomass (Grass only)								
April-2012	77	619	372	347	144	39	31	21
Jul-2012	68	592	260	232	126	49	25	26*
March-2013	72	1152	408	354	210	52	36	34*

*STDEV= standard deviation, CV=coefficient of variation, SE = standard error, * indicates that the data are not normally distributed according to Shapiro-Wilk normality test (p<0.05) (Royston, 1982).*

Prediction of leaf N concentrations

Overall the RF models combining grass and trees explained over 89% of leaf N concentrations compared to less than 89% for the grass only predictions (Table 2; Figure 1; 2). For the tree – grass modelling, the April and July 2012 models explained over 95% of leaf N concentrations, while the March 2013 explained only 89% of leaf N concentrations. For the grass modelling, the highest prediction was achieved in March 2013 (89%), compared to July and April 2012 (71% and 81% respectively, Table 2). Similar results were achieved by Ramoelo et al. (2012) using RapidEye imagery, demonstrating that a good prediction of leaf N can be achieved during peak productivity. Knox et al. (2011) also showed a possibility to estimate leaf N during the dry season but the estimation accuracy was relatively low, which is comparable with the results achieved in this study. The leaf N estimation accuracy of the

grass based on RMSE does not differ in wet and dry season. This may suggest the prediction power and robustness of RF for estimating vegetation parameters, though the developed models are season and data specific. Random forest is a nonparametric technique which does not have any assumption about the statistical distribution (Breiman, 2001; Mutanga et al., 2012).

Table 2: Performance of various data sets and vegetation indices to predict leaf N and biomass

Data sets	R²	RMSE	RRMSE (%)	P
Leaf N (Grass+Trees)				
April-2012	0.96	0.12	18	< 0.05
July-12	0.97	0.10	19	< 0.05
March 2013	0.89	0.26	33	< 0.05
Leaf N (Grass only)				
April-12	0.81	0.05	8	< 0.05
July-12	0.71	0.04	8	< 0.05
March 2013	0.89	0.08	12	< 0.05
Biomass (Grass only)				
April-12	0.90	59.32	17	< 0.05
July-12	0.91	45.22	19	< 0.05
March-13	0.84	105.61	30	< 0.05

*Relative RMSE ((RMSE/Observed Median)*100)*

Important selected variables for predicting leaf N using RF varied according to seasons for the combined tree – grass models (Figure 3). In April 2012, SR and NDVI were important variables. Interestingly, the red-edge based SR and NDVI were among the top 5 of highly ranked variables. Generally, the red-edge based SR was selected in the top 5 of important variables in all dates. MTCI2 which is red edge based, was selected in top 5 in March 2013 (grass+trees) and July 2012 (Grass only). Original bands B2 (480 nm) and B3 (545 nm) were also highly ranked in the top 5, in March 2013. Similar trends were observed in the estimation of leaf N for the grass only models, where red edge-based SR and NDVI were selected as important variables. Most of the selected bands (e.g. B2 and B4 (605 nm)) are chlorophyll absorption bands in the visible region of the spectrum. Inclusion of red-edge bands in vegetation indices was previously reported to improve the estimation of leaf N, using hyperspectral data (Gong et al., 2002), or satellite multispectral data, for instance with RapidEye images (Cho et al., 2013; Ramoelo et al., 2012). Red-edge bands are known to highly correlate with chlorophyll (Cho and Skidmore, 2006; Clevers et al., 2002) and are less sensitive to background effects (Zarco-Tejada et al., 2004). The basis for estimating leaf N using red-edge is that there is a positive relationship between chlorophyll and nitrogen in plants (Vos and Bom, 1993; Yoder and Pettigrew-Crosby, 1995).

Figure 1
Figure 2
Figure 3

Prediction of grass biomass

The RF models explained over 90% of biomass variation, with RMSE of 59.32 and 45.22 g/m^2, respectively in April and July 2012 (Table 2, Figure 4). The lowest RMSE was achieved in July 2012. In March 2013, the RF model explained 84% of biomass variation with the highest RMSE of 105.65 g/m^2, which is about 30% of the mean. Thus, grass biomass was predicted with a relatively lower accuracy in March 2013 compared to the other seasons. The estimation of biomass during the wet season has proved to be challenging, because of the saturation problems especially when using vegetation indices(Tucker, 1977). Figure 4 shows that the estimates of biomass are underestimated at the highest range of biomass in March 2013. Hyperspectral studies focussing on biomass estimation showed that narrow vegetation indices based on red-edge band solves the saturation problems (Mutanga and Skidmore, 2004b), however best results were obtained with narrow band red-edge (i.e. 1-5 nm) which are available with field or airborne hyperspectral data (Cho et al., 2007). Mutanga et al. (2012) demonstrated that biomass can be estimated on wetland species using a combination of vegetation indices derived from WorldView-2 and RF. Other few studies tested RF for retrieving vegetation parameters (Ismail and Mutanga, 2010), and showed RF is a robust methodology for feature or variable selection and prediction of biomass and leaf N. Nevertheless, Mutanga et al., (2012) argued that RF is sensitive to higher values, especially for biomass, and this could be associated with the saturation problem. This should be further explored. However, there is a significant improvement of biomass estimation using RF and WorldView-2 data. Details on challenges and limitations for estimating aboveground biomass using remote sensing has been outlined in published literature (Lu, 2006).

Figure 4

Variables of importance in the analysis showed that the red edge based SR was one of the top 5 for predicting biomass, in July 2012 and March 2013 (Figure 5). In March 2013, B1 (425 nm), B2 (480 nm), B7 (855 nm), and B8 (950 nm) are some of the bands ranked important for predicting biomass. The B1 and B2 were selected because they are sensitive to chlorophyll while B7 and B8 are sensitive to leaf structure, i.e. leaf area index (LAI) and biomass (Kumar et al., 2001). In July 2012, MTCI and B4 (605 nm) were among the top 5 which is indicator of some relationship between leaf N (related to chlorophyll) and biomass (Plummer, 1988). Similar results were observed in April 2012, where vegetation indices dominated the top 5 of the highly ranked variables.

Figure 5

Seasonal estimation of leaf N and biomass using RF

The leaf N and biomass results are different from the various seasons. Phenology is a major driver of this outcome (Wenjiang et al., 2004), especially when using vegetation indices which depend on the vegetation greenness. The results showed that leaf N models are more driven by the tree leaf N values in winter (July) than by the grass N. This was also evidenced by the higher prediction capability when the grass and tree data sets were combined. Most trees especially evergreen ones tap water from underground during dry season, sustaining greenness throughout the season, than the deciduous trees which shed leaves during dry season.

The larger prevalence of high leaf N values is evident in March 2013 (Figure 6) as compared to July 2012 (Figure 7). The high leaf N values in March 2013 depicted by red colour are

mostly trees and shrubs and are evident in the riparian zones and along the fence extending from north to south. The major driver of the fence-line distribution of leaf N in March 2013 is geology (See Figure 6). From Sabi Sands and after the north-south fence-line, geology is mainly gabbro and settlement areas are dominated by granite. While in July, high leaf N values are concentrated in the riparian or riverine areas which are mainly associated with trees than grass. The other factor influencing leaf N distribution patterns is the greenness of vegetation which could be linked to rainfall (Pickett et al., 2003). The riverine areas exhibit high leaf N values in both March 2013 and July 2012, which could be associated with the deposited soil nutrients from the crest (Venter et al., 2003) and also availability of water in the rivers during winter season. In addition to the riverine areas, bottomlands are also associated with high leaf N values, which show the importance of topography on leaf N distribution (Scholes et al., 2003; Venter et al., 2003).

Figure 6
Figure 7

Over-land difference in grass biomass is evident between the wet season in March 2013 (Figure 8) and the dry season in July 2012 (Figure 9). As indicated above, distribution of biomass might be influenced by rainfall and phenology cycles (Pickett et al. 2003). These areas receive high rainfall in summer (wet season) and limited or none in winter (dry season). The protected areas (Sabi Sands) generally have high grass biomass as compared to the communal areas, in March 2013. In July, low biomass found in the protected and communal areas could be associated to fire and overgrazing. In the protected areas, high biomass values are concentrated in the riparian areas; this could be linked to the availability of water in dry season along the riparian areas. Geology is also a factor influencing the distribution of biomass as depicted in Figure 8. Gabbro areas have high grass biomass as compared to the surrounding areas dominated by granite. Gabbro is nutrient and clay-rich, and is not easily weathered as compared to granite (Venter et al., 2003). Clay-rich soils have high soil water holding capacity, and influence higher grass production (Venter et al. 2003).

Figure 8
Figure 9

The relationship between the predicted biomass and leaf N is poor in both wet and dry season, with R^2 of 0.23 and 0.04, respectively. The Spearman correlation and cross-variogram analysis showed that biomass and leaf N are poorly and negatively related (*rho* = -0.46) especially for March 2013, and do not co-vary with a negative semivariance (Figure 10; 11), and no relationship (*rho*=0.12) for July 2012. The Exponential (Figure 10) and Gaussian (Figure 11) models were successfully fitted to the variogram, and used for cross-variogram analysis, for July 2012 and Mach 2013, respectively. The interaction of the various factors influencing the distribution of leaf N and biomass presents the complexity of the savanna ecosystems. These complexities of the savanna ecosystem are scale dependent. For example, biomass and leaf N distribution in relation to topography is not always linear. Crest areas which are known to have low biomass and low nutrients, can have high biomass for species such as *Hyperthelia dissoluta*, with low nutrient content. Therefore, at small scale or lower extent, the relationship between biomass and leaf N might not be good, but wider areas or larger extent, the relationship might be good. This issue of scales requires further investigation.

Figure 10
Figure 11

The key question investigated in this study was "can new high spatial multispectral imagery with red-edge bands such as WorldView-2 be used to develop an approach to move towards regional assessment and monitoring of leaf N and biomass as indicators of rangeland quality and quantity"? The study demonstrated that leaf N and biomass can be estimated using vegetation indices in dry and wet season. Season specific analysis showed that wet season model (March model) performed better than the dry season one (July model) leaf N. This study fulfils what was achieved by Ramoelo et al., 2012, Cho et al., (2013), Skidmore et al. (2010) as well as Knox et al. 2011. Other studies using hyperspectral data showed that narrow band absorption features in the shortwave infrared (SWIR) region are sensitive to leaf N and protein, and are crucial for estimation of leaf N in dry or green vegetation (Elvidge, 1990; Kokaly, 2001). Protein occurs in the form of ribulose 1-5-biphosphate carboxylase-oxugenase (rubisco) which is the most N bearing compounds in green leaves (Elvidge, 1990; Kokaly et al., 2009; Kumar et al., 2001). Protein contains about 30 to 50% of leaf N in fresh leaves (Elvidge, 1990). Protein and N features are centered on 1020nm, 1510nm, 2690nm, 1730nm, 1940nm, 1980nm, 2060nm, 2180nm, 2240nm, 2300nm, 2350nm (Curran, 1989; Kokaly, 2001; Kumar et al., 2001). Furthermore, in literature, assessment of dead vegetation parameters (e.g. biomass) were achieved by field and imaging spectroscopy using an index such as Cellulose Absorption Index (CAI) (Nagler et al., 2003). This was rarely studied using multispectral remote sensing (Xu et al., 2014), because of the paucity of narrow band SWIR. Similar to this study, Xu et al. 2014 demonstrated that there is a relationship between dead material and broadband multispectral vegetation indices such as NDVI. Several studies used vegetation indices for mapping vegetation dynamics in dry environments such as desert (Anyamba and Tucker, 2005). In this study, both leaf N and biomass were estimated during dry season with acceptable accuracy. There is a greater potential to further explore remote sensing of senescence or dead vegetation using multispectral vegetation indices, as it is crucial for grazing purposes.

Conclusions

Prediction of leaf N and biomass across season is plausible using RF and WorldView-2 imagery with the red edge band capability. The assessment and monitoring of leaf N and biomass as indicators of rangeland quality and quantity can be done using RF and WorldView-2 based indices. Dry season estimation of leaf N and biomass is possible using a combination of vegetation indices and reflectance from WorldView-2 bands and there is a potential to further explore remote sensing of senescence vegetation. This study provides a basis for rangeland monitoring and assessment at even larger scales using data sets like the upcoming Sentinel-2 with the red edge capability across dry and wet seasons. This research provides information which could be crucial for decision makers on planning and management of rangeland systems e.g. for understanding population dynamics, disease transmission, spatial zoning for grazing camps and determining carrying capacity for herbivores.

Acknowledgement

References

Adam, E.M., 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. International Journal of Remote Sensing 33, 552-569.

Anyamba, A., Tucker, C.J., 2005. Analysis of Sahelian vegetation dynamics using NOAA-AVHRR NDVI data from 1981–2003. Journal of Arid Environments 63, 596-614.

Bengis, R.G., Grant, C.C., De Vos, V., 2003. Wildlife Diseases and Veterinary Controls: A Savanna Ecosystem Perspective, in: Du Toit, J.T., Rogers, K.H., Biggs, H.C. (Eds.), The Kruger Experience: Ecology and Management of Savanna Heterogeneity. Island Press, London.

Bivand, R.S., Pebesma, E.J., Gomez-Rubio, V., 2008. Applied Spatial Data Analysis. Springer.

Bothma, J.d.P., 1989. Game Ranch Management. J.L. van Schaik (PTY) Ltd, Pretoria.

Breiman, L., 2001. Random Forests. Machine Learning 45, 5-32.

Chaminuka, P., McCrindle, C.M.E., Udo, H., 2011. Cattle farming at the wildlife/livestock interface: assessment of costs and benefits Adjacent to Kruger National Park, South Africa. Society and Natural Resources DOI:10.1080/08941920.2011.5804171-16.

Cho, M., Skidmore, A.K., Corsi, F., van Wieren, S., Sobhan, I., 2007. Estimation of green grass/herb biomass from airborne hyperspectral imagery using spectral indices and partial least square regressions. International Journal of Applied Earth Observation and Geoinformation 9, 414-424.

Cho, M.A., Ramoelo, A., Debba, P., Mutanga, O., Mathieu, R., van Deventer, H., Ndlovu, N., 2013. Assessing the effects of subtropical forest fragmentation on leaf nitrogen distribution using remote sensing data. Landscape Ecology 28, 1479-1491.

Cho, M.A., Skidmore, A.K., 2006. A new technique for extracting the red edge position from hyperspectral data: The linear extrapolation method. Remote Sensing of Environment 101, 181-193.

Clevers, J.G.P.W., De Jong, S.M., Epema, G.F., Van Der Meer, F.D., Bakker, W.H., Skidmore, A.K., Scholte, K.H., 2002. Derivation of the red edge index using the MERIS standard band setting. International Journal of Remote Sensing 23, 3169 - 3184.

Clifton, K.E., Bradbury, J.W., Vehrencamp, S.L., 1994. The fine-scale mapping of grassland protein densities. Grass Forage Sci. 49, 1-8.

Curran, P.J., 1989. Remote sensing of foliar chemistry. Remote Sensing of Environment 30, 271-278.

Darvishzadeh, R., Skidmore, A., Schlerf, M., Atzberger, C., Corsi, F., Cho, M., 2008. LAI and chlorophyll estimation for a heterogeneous grassland using hyperspectral measurements. ISPRS Journal of Photogrammetry and Remote Sensing 63, 409-426.

Dash, J., Curran, P.J., 2004. The MERIS terrestrial chlorophyll index. International Journal of Remote Sensing 25, 5403 - 5413.

Devlin, R.M., 1975. Plant Physiology. Litton Educational Publishing, Inc., New York.

Diaz-Uriarte, R., Alvarez de Andres, S., 2006. Gene selection and classification of microarray data using random forest. BMC Bioinformatics 7, 3.

Dijkshoorn, K., van Engelen, V., Huting, J., 2008. Global Assessment of Land Degradation: Soil and Landform propertes for Land Degradation Assessment in Drylands (LADA) partner countries (Argentina, China, Cuba, Senegal and the Gambia, South Africa and Tunisia), ISRIC Report 2008/06 and GLADA Report 2008/03. ISRIC-Word Soil Information ; FAO-Food and Agriculture Organization of the United Nations, Wageningen.

Drent, R.H., Prins, H.H.T., 1987. The herbivore as prisoner of its food supply, in: Andel, J.V., Bakker, J. (Eds.), Disturbance in Grasslands: Species and Population Responses. Dr. W. Junk Publishing Company, Dordrecht, pp. 133-149.

Efron, B., Gong, G., 1983. A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation. The American Statistician 37, 36-48.

Elvidge, C.D., 1990. Visible and near infrared reflectance characteristics of dry plant materials. International Journal of Remote Sensing 11, 1775-1795.

Gong, P., Pu, R., Heald, R.C., 2002. Analysis of in situ hyperspectral data for nutrient estimation of giant sequoia. International Journal of Remote Sensing 23, 1827 - 1850.

Grant, C.C., Scholes, M.C., 2006. The importance of nutrient hot-spots in the conservation and management of large wild mammalian herbivores in semi-arid savannas. Biological Conservation 130, 426-437.

Ham, J., Chen, Y., Crawford, M.M., Ghosh, J., 2005. Investigation of the random forest framework for classification of hyperspectral data. IEEE Transactions on Geoscience and Remote Sensing 43, 492-501.

Hollander, M., Wolfe, D.A., 1973. Nonparametric Statistical Methods. John Wiley & Sons, New York.

Huang, Z., Turner, B.J., Dury, S.J., Wallis, I.R., Foley, W.J., 2004. Estimating foliage nitrogen concentration from HYMAP data using continuum removal analysis. Remote Sensing of Environment 93, 18-29.

Ismail, R., Mutanga, O., 2010. A comparison of regression tree ensembles: Predicting Sirex noctilio induced water stress in Pinus patula forests of KwaZulu-Natal, South Africa. International Journal of Applied Earth Observation and Geoinformation 12, Supplement 1, S45-S51.

Jordan, C.F., 1969. Derivation of leaf area index from quality of light on the floor. Ecology 50, 663-666.

Knox, N.M., Skidmore, A.K., Prins, H.H.T., Asner, G.P., van der Werff, H.M.A., de Boer, W.F., van der Waal, C., de Knegt, H.J., Kohi, E.M., Slotow, R., Grant, R.C., 2011. Dry season mapping of savanna forage quality, using the hyperspectral Carnegie Airborne Observatory sensor. Remote Sensing of Environment 115, 1478-1488.

Kokaly, R.F., 2001. Investigating a physical basis for spectroscopic estimates of leaf nitrogen concentration. Remote Sensing of Environment 75, 153-161.

Kokaly, R.F., Asner, G.P., Ollinger, S.V., Martin, M.E., Wessman, C.A., 2009. Characterizing canopy biochemistry from imaging spectroscopy and its application to ecosystem studies. Remote Sensing of Environment 113, S78-S91.

Kumar, L., Schmidt, K.S., Dury, S., Skidmore, A.K., 2001. Imaging Spectroscopy and Vegetation Science, in: Van Der Meer, F.D., De Jong, S.M. (Eds.), Image Spectroscopy. Kluver Academc Publishers, Dordrecht, pp. 111-154.

Liaw, A., Wiener, M., 2002. Classification and regression by Random Forest. R News 2.

Lu, D., 2006. The potential and challenge of remote sensing-based biomass estimation. International Journal of Remote Sensing 27, 1297-1328.

Ludwig, F., De Kroon, H., Prins, H.H.T., 2008. Impacts of savanna trees on forage quality for large African herbivore. Oecologia 155, 487-496.

Ludwig, F., de Kroon, H., Prins, H.H.T., Berendse, F., 2001. Effects of nutrients and shade on tree-grass interactions in an east African savanna. Journal of Vegetation Science 12, 579-588.

McNaughton, S.J., 1988. Mineral nutrition and spatial concentrations of African ungulates. Nature 334, 343-345.

McNaughton, S.J., 1990. Mineral nutrition and seasonal movements of African migratory ungulates. Nature 345, 613-615.

Mutanga, O., Adam, E., Cho, M.A., 2012. High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm. International Journal of Applied Earth Observation and Geoinformation 18, 399-406.

Mutanga, O., Skidmore, A.K., 2004a. Integrating imaging spectroscopy and neural networks to map grass quality in the Kruger National Park, South Africa. Remote Sensing of Environment 90, 104-115.

Mutanga, O., Skidmore, A.K., 2004b. Narrow band vegetation indices overcome the saturation problem in biomass estimation. International Journal of Remote Sensing 25, 3999 - 4014.

Nagler, P.L., Inoue, Y., Glenn, E.P., Russ, A.L., Daughtry, C.S.T., 2003. Cellulose absorption index (CAI) to quantify mixed soil-plant litter scenes. Remote Sensing of Environment 87, 310-325.

Owen-Smith, N., Danckwerts, J., E., 1997. Herbivory, in: Cowling, R.M., Richardson, D.M., Pierce, S.M. (Eds.), Vegetation of Southern Africa. Cambridge University Press, Cambridge.

Pickett, S.T.A., Gadenasso, M.L., Benning, T.L., 2003. Biotic and Abiotic Variability as Key Determinants of Savanna Heterogeneity at Spatiotemporal Scales, in: Du Toit, J.T., Rogers, K.H., Biggs, H.C. (Eds.), The Kruger Experience: Ecology and Management of Savanna Heterogeneity. Island Press, London, pp. 22-40.

Plummer, S.E., 1988. Exploring the relationships between leaf nitrogen content, biomass and the near-infrared/red reflectance ratio. International Journal of Remote Sensing 9, 177-183.

Ramoelo, A., Skidmore, A.K., Cho, M.A., Mathieu, R., Heitkönig, I.M.A., Dudeni-Tlhone, N., Schlerf, M., Prins, H.H.T., 2013. Non-linear partial least square regression increases the estimation accuracy of grass nitrogen and phosphorus using in situ hyperspectral and environmental data. ISPRS Journal of Photogrammetry and Remote Sensing 82, 27-40.

Ramoelo, A., Skidmore, A.K., Cho, M.A., Schlerf, M., Mathieu, R., Heitkönig, I.M.A., 2012. Regional estimation of savanna grass nitrogen using the red-edge band of the spaceborne RapidEye sensor. International Journal of Applied Earth Observation and Geoinformation 19, 151-162.

Ren, H., Zhou, G., 2012. Estimating senesced biomass of desert steppe in Inner Mongolia using field spectroscopy. Agricultural and Forest Meteorology 161, 66-71.

Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W., Harlan, J.C., 1974. Monitoring the vernal advancement and retrogradation of natural vegetation, NASA/GSFC, Type III Final Report, M.D. Greenbelt, 371.

Royston, P., 1982. Algorithm AS 181: the W test for normality. Applied Statistics 31.

Scholes, R.J., Bond, W.J., Eckhardt, H.C., 2003. Vegetation Dynamics in the Kruger Ecosystem, in: Du Toit, J.T., Rogers, K.H., Biggs, H.C. (Eds.), The Kruger Experience : Ecology and Management of Savanna Heterogeneity. Island Press, London.

Shackleton, S.E., Shackleton, C.M., Netshiluvhi, P.R., Geach, B.S., Ballance, A., Fairbanks, D.H.K., 2002. Use patterns and value of savanna resources in three rural villages in South Africa. Economic Botany 56, 130-146.

Skidmore, A.K., Ferwerda, J.G., Mutanga, O., Van Wieren, S.E., Peel, M., Grant, R.C., Prins, H.H.T., Balcik, F.B., Venus, V., 2010. Forage quality of savannas -- Simultaneously mapping foliar protein and polyphenols for trees and grass using hyperspectral imagery. Remote Sensing of Environment 114, 64-72.

Treydte, A.C., Heitkönig, I.M.A., Prins, H.H.T., Ludwig, F., 2007. Trees improve grass quality for herbivores in African savannas. Perspectives in Plant Ecology, Evolution and Systematics 8, 197-205.

Tucker, C.J., 1977. Asymptotic nature of grass canopy spectral reflectance. Applied Optics 16.

Venter, F.J., Scholes, R.J., Eckhardt, H.C., 2003. Abiotic template and its associated vegetation pattern, in: Du Toit, J.T., Kevin, H.R., Biggs, H.C. (Eds.), The Kruger Experience: Ecology and Management of Savanna Heterogeneity. The Island Press, London.

Vos, J., Bom, M., 1993. Hand-held chlorophyll meter: a promising tool to assess the nitrogen status of potato foliage. Potato Research 36, 301-308.

Wenjiang, H., Jihua, W., Zhijie, W., Jiang, Z., Liangyun, L., Jindi, W., 2004. Inversion of foliar biochemical parameters at various physiological stages and grain quality indicators of winter wheat with canopy reflectance. International Journal of Remote Sensing 25, 2409-2419.

Xu, D., Guo, X., Li, Z., Yang, X., Yin, H., 2014. Measuring the dead component of mixed grassland with Landsat imagery. Remote Sensing of Environment 142, 33-43.

Yoder, B.J., Pettigrew-Crosby, R.E., 1995. Predicting nitrogen and chlorophyll content and concentrations from reflectance spectra (400-2500 nm) at leaf and canopy scales. Remote Sensing of Environment 53, 199-211.

Zarco-Tejada, P.J., Miller, J.R., Morales, A., Berjon, A., Aguera, J., 2004. Hyperspectral indices and model simulation for chlorophyll estimation in open-canopy tree crops. Remote Sensing of Environment 90, 463-476.