Estimation of canopy nitrogen concentration across C3 and C4 grasslands using WorldView-2 multispectral data

C. Adjorlolo, O. Mutanga, and M.A. Cho

***Abstract*—This paper assesses the potential of multispectral data attained using WorldView-2 (WV2) satellite to estimate and map the variability in canopy nitrogen (N) concentration across C3 and C4 grasslands. The WV2 satellite image was acquired for the Cathedral Peak region of the Drakensberg Mountain range, South Africa. The Random Forest (RF) regression algorithm was used to develop a relationship between two-band normalized ratio indices (RIs), including the Normalized Difference Vegetation Index (NDVI), computed from the WV2 image data and N concentration. The RF-based variable importance scores calculated using the training dataset (n = 150) showed that the RI computed involving the costal-blue (400–450 nm) and yellow (585–625 nm) band is the most important, when predicting canopy N concentration in the area. Using the validation dataset (n = 64), the RF explained 71% of the variation, with a Nash–Sutcliffe efficiency (NSE) = 0.68, in predicting N across the C3 grass, *Festuca costata,* and C4 grasses, *Themeda triandra* and *Rendlia altera* grasslands. Overall, results indicate that predictions of canopy N using new multispectral data with unique band setting, such as WV2 spectra are possible for C3 and C4 grasslands.**

Index Terms – Grass nitrogen content, worldview-2, random forests, remote sensing.

I. Introduction

Vegetation carbon and nitrogen (N) content in rangeland ecosystems have been strongly linked through a number of critical ecological processes, such as net primary production of plant functional types, litter decomposition, and rates of photosynthesis [1, 2]. The net primary production of quality biomass is one of the outcomes of complex interaction between vegetation and environment, which is one of the key factors of rangeland condition. Such interaction is widely considered in global climate change research that focuses on studying the impacts of climatic variables (e.g. elevated CO_2 and temperatures) on growth responses of C3 and C4 types of grass species [3]. The C3 or C4 denotes an initial photosynthetic product of 3-carbon and 4-carbon compounds, respectively. Studies have highlighted that the rates of assimilation or conversion of macro nutrients under changing environmental or climatic conditions differ between C3 and C4 grass species [3-8]. For example, these two groups of grasses generally differ, with the C3 grass type having higher concentration of N, but lower carbon (C)/N ratio, compared to C4 grass type [9].

At landscape scale, C3 and C4 grass forage nutrients can vary in response to multiple factors such as precipitation, fire, herbivory and site-specific edaphic conditions [3, 10]. These factors are inextricably linked, creating a patchy distribution of forage nutrients across landscapes composed of diverse grass species [11, 12]. Currently, the prediction and mapping of canopy nutrient concentration, specifically, across C3 and C4 grasslands is lacking due to the fine-scale nature of the variation and large abundance of grass species [13]. However, the estimation and mapping of N concentration across C3 and C4 grasslands should provide insight into their carbon and N cycles and, therefore, be a useful indicator into models of plant functional type dynamics and ecosystem productivity [14].

Remote sensing applications have proved to be a source of proximal data for estimating several canopy attributes relating to biophysical, physiological or biochemical characteristics of vegetation [14-16]. The remote sensing hyperspectral systems, which acquire reflectance data in hundreds of narrow (< 10 nm), contiguous spectral channels, over the visible, near-infrared and shortwave-infrared portions of the electromagnetic spectrum (400–2500 nm) have been more successful, compared to traditional broadband (>100 nm) multispectral sensors [17]. The major challenge is that although hyperspectral systems provide a continuous spectrum for each pixel of a scene, such data are high in dimension, not easily accessible and can be relatively costly [18]. In this regard, new generation, space-borne multispectral sensors, such as DigitalGlobe's WorldView-2 (WV2) satellite with unique band settings offer new opportunities for characterizing the biophysical and biochemical attributes of vegetation [19, 20].

The normalized difference vegetation index (NDVI) based on a combination of bands from the red and infrared wavelengths or normalized ratio indices (RIs), often calculated from two-band combination, have been used to improve the ability to predict the concentration per unit area of biochemicals at leaf or canopy scales [21-23]. The NDVI

Manuscript submitted on 30 October 2013

Authors of this study are thankful of the University of KwaZulu-Natal for providing the necessary support. Additional support was provided by the National Research Foundation (NRF), the KwaZulu-Natal Department of Agriculture and Environmental Affairs (KZN DAE) and the Ezemvelo KZN Wildlife.

The authors are with the South African National Space Agency (SANSA), Pretoria 0087; School of Agriculture, Earth and Environmental Sciences, University of KwaZulu-Natal, P. Bag X01, Scottsville 3209, Pietermaritzburg, South Africa, and the Natural Resources and the Environment, Council for Scientific and Industrial Research (CSIR), Pretoria 0001, South Africa, respectively. Corresponding author e-mail: cadjorlolo@sansa.org.za

is sensitive to the difference between high reflectance in the near infrared and low reflectance in the red region to predict biophysical parameters. The difference in reflectance spectra is strongly related to increased absorption in the red region by chlorophyll for vegetation with greater chlorophyll concentration, combined with increased near infrared scattering by vegetation structural components, for vegetation with high biomass. Plant N constitutes a significant part of the chlorophyll concentration in the leaf canopy. Nitrogen status in plant canopy significantly affects leaf area, leaf chlorophyll content and photosynthetic rate, often altering biomass production [15, 24]. Zhao et al. [24] found that leaf N and chlorophyll concentrations were linearly correlated with reflectance ratios of 405/715 nm and 1075/735 nm for C4 crop, sorghum (*Sorghum bicolor*). Hansena and Schjoerring [15] illustrated that RIs calculated on wavebands centered at 440, 447, 459 and 573 nm wavelengths were useful for predicting percentage N concentration for wheat crops. These bands centers are within the spectral range for WV2 sensor (i.e. coastal blue, blue, green and yellow channels).

In order to predict variability in biochemical or patterns at landscape scale, it is critical that vegetation indices should be universal and species-independent, accounting for the species-specific differences in reflectance, scattering and internal or canopy structure [23, 25]. Several data analysis approaches, including those based on parametric techniques (e.g. simple stepwise linear regression and partial least squares) have been commonly used to combine the information content of remotely sensed vegetation variables [21]. The major challenge is that such parametric techniques have often yielded limited results, largely because of their requirements to satisfy some statistical conditions or to assume normal distribution of the input dataset. On the other hand, advanced techniques based on machine learning algorithms: artificial neural networks [26], support vector machines [27] and RF [19], which make no assumption of the input data distribution have increasingly offered a better capability to analyze remote sensing data [28, 29]. From this background, the current study assesses the capability of combining the information content of VIs computed from the WV2 image data for estimation of canopy N concentration. The paper further assesses the utility of the RF regression in mapping variation in N concentration across a landscape composed of C3 and C4 grass species.

II. METHODS

A. The study area

The study area is located in the Cathedral Peak region (uKhahlamba World Heritage Site) of the Drakensberg Mountain range in KwaZulu-Natal province, South Africa. The specific sampling site (Fig. 1) is underlain by relatively homogeneous Drakensberg basalt form of the Stormberg series. Patches of the C3 and C4 type of grasslands on crests and in open valleys were sampled, across the Montane and sub-alpine vegetation belts (1280–2865 m above mean sea level). The sub-alpine belt has been described as *Themeda-Festuca*, sub-alpine grassland [30] and is composed of the C3 and C4 type of grass species. It is further divided into two grassland communities known as the *Themeda triandra* (red grass) and *Festuca costata* (tussock fescue) sub-alpine grasslands. Notably, there are strong associations of *T. triandra* and *F. costata* occurrence on warm, northerly and cool, southerly slopes, respectively [31, 32]. The study area is also characterized by high rainfall (> 800 mm per annum). Misty conditions, low temperatures, seasonal frosts, snow, fire, herbivory and topographic factors play an important role in influencing the composition and functioning of C3 and C4 grasses in the area [33].

The C4 grassland communities consist mainly of the *T. triandra* and *Rendlia altera* (Toothbrush grass) species, while the C3 community consists mainly of unpalatable *F. costata* and its co-existing, coarse C3 species, *F. caprina*, *Merxmuellera macowanii* and *Pteridium aquilinum* (Bracken fern). The *F. costata*, *F. caprina* and *M. macowanii* remain evergreen throughout the dry (winter) season. They normally grow as large tussocks that are taller than the surrounding C4 grasslands. They also grow in an aggregated pattern, which makes them easily distinguishable in the field.

The dominant C4 grassland communities can make up 70%, or more, of the composition in C4 community or a patch. The *R. altera* is characterized by relatively narrow leaves, which tend to fold and break-up into fibres in mature plants. This results in low leaf production and in a relatively low canopy density. The *T. triandra* forms dense and tufted stands that are distinctly red-coloured, when they are fully mature.

The study area consists of catchments which reflect some dominant species or vegetation communities, wetness and topographic gradients, and therefore a variation in foliar N concentration. The target C3 and C4 communities or patches were already classified by previous studies conducted in the area [34, 35]. Thus, areas of interest sampled for the current study are consistent with the spatial distribution of the dominant C3 and C4 grasses under study.

B. Plant sample collection

The field campaign was conducted to coincide with the date (28 April 2011) of the WV2 image acquisition. All target plants were at a mature state and thus full canopy cover. The stratified random sampling approach was adopted to place sample plots in each of the target grassland patches. A total of 64 plots were sampled for *F. costata*, 49 plots for *R. altera* and 101 plots for *T. triandra* grassland. The targets, namely the mature grass species in a 1×1 m sub-plots were clipped, just above the ground. These plots were randomly established by throwing a 1 m^2 quadrat, about three times, within a larger 10 × 10 m vegetation plots. All the plots sampled had 60% or more of the candidate grass species. The above-ground grass biomass was

harvested and weighed *in situ*, and then transported in brown paper bags, placed in a cooler box, to the feed lab within two to four hours of harvesting.

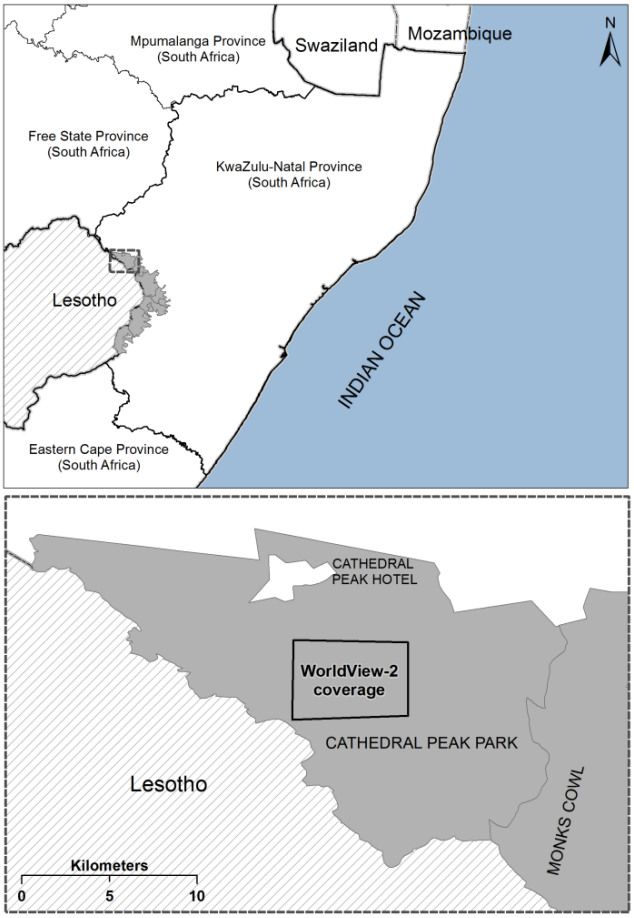

Fig. 1. Location of the Cathedral Peak study area

C. Chemical analysis

The harvested grass samples for each target species were dried at 70°C for 48 hours in an oven and subsequently weighed. The dried samples were then mill-crushed to about 1 mm. These dried and crushed samples were processed for complete feed analyses, including the percentage N concentration, calculated on a 100% dry-matter (DM) basis for samples. The Kjeldahl procedure was used to determine the concentration of nitrogen. The prior digestion of the samples was done in a mixture of sulphuric acid, selenium and salicylic acid [36]. The samples were then colorimetrically measured, using a continuous flow analyser (SKALAR SAN plus).The resultant data were pooled for use in the statistical analysis phase.

D. Image acquisition and pre-processing

The WV2 multispectral image used in this study comprises eight (8) variable spectral channels, spanning the wavelength range: 400–450 nm (B1-coastal blue), 450–510 nm (B2-blue), 510–581 nm (B3- green), 585–625 nm (B4-yellow), 630–690 nm (B5-red), 705–745 nm (B6-red edge), 770–895 nm (B7-near infrared- 1), and 860–1040 nm (B8-near infrared-2). The spatial resolution of the 8 bands was 2 m, covering an area of 25 km^2. The data were delivered at product level LV3D (DigitalGlobe, Longmont, Colorado, USA). That means the image had been radiometrically corrected and orthorectified by the supplier. The mean in-track, cross-track and off-nadir view angles were -3.1°, 14.9° and 15.3°, respectively. The WV2 satellite's mean azimuth and elevation angles were 112.1° and 72.9°, respectively, and the sun azimuth and sun elevation angles were 32.6° and 40.8°, respectively. The registration mean error of the image was 0.15 m. The digital numbers were converted to WV2 radiance, using the absolute radiometric calibration factors and effective bandwidths for each band, using ENVI 4.8 software.

Considering that the WV2 image was acquired for a mountainous landscape, it was necessary to process the data, in order to correct for the topographic effects of the rugged and undulating terrain, and the Sun's geometry. Topographic corrections were calculated with the only available 20 m digital elevation module (DEM). However, it turned out that the 20 m DEM was insufficient in correcting the topographic effects on the WV2 dataset, for the landscape under investigation. This was the case, even when the 20 m DEM was re-sampled step by step from 20 m to 10 m, and 5 m, to match the WV2 spatial resolution. The output images were inaccurate, with a substantial positional shift between the DEM and the WV2 dataset, showing a chessboard pattern of the output image. Therefore, topographic correction was excluded from the analysis, in order not to introduce an additional error to the reflectance values. In this regard, steep slopes and sheltered areas were masked from all analysis.

E. Vegetation indices

Using the 8 bands of WV2 image data, a total of $n \times n = 56$ normalized ratio indices were computed from all possible two-band combinations. Normalized ratio indices or RIs combine the vegetation reflectance spectra present between two or more wavebands to enhance the spectral contribution from the vegetation canopy while minimizing perturbing factors such as substrate water and soil color changes or atmospheric effects. The differences in canopy reflectance associated with changes in plant N concentration are influenced by changes in vegetation cover, leaf area index (LAI) and aboveground biomass [15, 22, 23, 37]. The current study sought after exploring the optimal band combinations for predicting nitrogen content in samples and assesses the empirical relationships for different C3 and C4 grass species. The normalized ratio indices were calculated for all WV2 wavebands to investigate their usefulness for estimating biophysical and biochemical properties of vegetation, including N concentration. Wavelengths of typical absorption peaks for N include: 1020 nm, 1510 nm, 1730 nm, 1980 nm , 2060 nm, 2130 nm, 2180 nm, 2240 nm, and 2300 nm [38, 39]. Given the spectral coverage for WV2,

the computation of classic N indices such as the Normalized Difference Nitrogen Index or NDNI [40] was not possible. Hence the multispectral 8 bands of WV2 allowed a computation of the RI as follows:

$$\mathrm{RI} = \frac{R_{(i,n)} - R_{(j,n)}}{R_{(i,n)} + R_{(j,n)}} \quad [1]$$

where $R_{(i,n)}$ and $R_{(j,n)}$ are the reflectance of any two bands for spectral sample, *n*. The resultant normalized ratio indices derived from the input WV2 bands were then assessed to find the coefficient of determination (R^2) between these indices and the N content of the samples and compared between C3/C4 species. This process yielded symmetrical R^2 values, for each pair of bands, computed using all possible combination of the shorter and longer wavelength bands. Therefore, only n = 28 variables were used in the RF regression analysis. All the R^2 values were plotted in a histogram plot which revealed a characteristic high and low pattern of R^2 values, indicating a number of ''hot spots'' with relatively high correlation coefficients.

F. Collecting image spectra

Pixels representing the dominant grasslands were used to estimate the minimum and maximum NDVI threshold values for the target grass species. Next, a region-of-interest map was created, using the NDVI variable. This was tested against very high spatial resolution (50 cm) WV2 panchromatic image data. Subsequently, the region-of-interest map was used to subset the preferred grassland areas, and pixels representing other vegetated surfaces or patches of bare soils (road and eroded surfaces) were masked out, using the mask tool in the ENVI4.8 software. A 1×1 pixel (4 m^2) corresponding to the centre coordinate of each sampled plot (1×1 m) locations was used to extract image spectra for all plots. The plots were located in homogenous sites and created around the centre co-ordinate.

III. DATA ANALYSIS

A. Random forest regression and variable importance

The RF [41] utilizes bootstrap samples with no replacement to construct a large number of regression trees. The RF bootstraps and sub-samples the training data (i.e. about 66% randomly and independently-selected samples) and then combines the predictions from the resulting ensemble of regression trees. These assign each RI variable to a response (N) value, based on the maximum number of votes that the value receives from the collection of all trees [41]. About one-third of the data, which were not included in the bootstrapped training sample, called out-of-bag (OOB) cases were used to evaluate the final model. The OOB samples offer an unbiased estimates of the prediction error [19]. The RF algorithm is easy to implement because only few parameters are tuned: (i) the number of regression trees to grow (*ntree*) and (ii) the bootstrapped number of predictors to split at each regression tree node (*mtry*). In the current study, the default *mtry* value, in the case of RF regression analysis, was calculated as *p*/3, whereby *p* is the number of predictors [41].

The RF algorithm offers variable importance (VIP) scores in its computation, providing researchers with valuable insights to explore the effect of each predictor variable on the response variable. The VIP measure was used to assess the contribution made by each of the WV2-derived RIs in predicting canopy N concentration. The assessment of the VIP is an important step for ranking the predictor variables before the final model estimation. The VIP also helps to interpret data and understand the underlying interacting processes. For each tree of the forest, the VIP score of a predictor variable was calculated as the normalized increase in mean square error of a tree when the observed values of this variable are randomly permuted. The difference in prediction accuracy prior and after permuting a variable in the OOB samples was used to calculate VIP score of a variable [42]. In this respect, the number of observations predicted correctly, decreases substantially if the permuted variable is strongly associated with the response variable. A detailed account of the random forest-based variable importance can be found in [43, 44].

B. Random forest-based forward variable selection

The RF algorithm not only yields very high regression accuracy but it is also a novel technique of modelling complex interactions among predictor variables and it is suitable for assessing variable importance and variable selection. The two main objectives of variable selection process are: 1) to identify key predictors highly correlated to the dependent variable for interpretation purpose; and 2) to identify a subset of predictors sufficient to fit a good prediction model of the dependent variable[45]. Application of the RF variable selection technique for both classification [33] and regression [46] analyses of remote sensing data has been demonstrated by several studies. Adjorlolo et al. [47] showed that the variables that RF identified as most important for classifying selected C3 and C4 species coincided with expectations based on the literature [48].

Using the RF ranked variable importance, a greedy fast forward variable selection (FvS) procedure was implemented [33]. The FvS procedure involved iteratively fitting the RF model on the training data and at each iteration building a new model by adding the predictor variable with highest importance. For each stage of the FvS process, the average OOB error was computed in order to assess the accuracy of the FvS model iterations. The optimum subset of RIs was determined using the iteration with the lowest OOB error.

C. Model validation

The performance of the models was measured by calculating the coefficient of determination (R^2), root mean square error (RMSE) and Nash–Sutcliffe efficiency (NSE).

The RMSE (Equation 2) provides a direct estimate of the model error expressed in the original measurement unit [21] whilst the NSE (Equation 3) index validates the model efficiency by assessing the relative magnitude of the residues compared to the variance in measured N concentration [49].

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y} - y)^2}{n}} \quad [2]$$

$$NSE = 1 - \left(\frac{\sum_{i=1}^{n} (\hat{y} - y)^2}{\sum_{i=1}^{n} (\hat{y} - \bar{y})^2} \right) \quad [3]$$

where ŷ indicates predicted N and y was the measured N value, ȳ represents the mean of measured N values and *n* is the number of samples.

IV. RESULTS AND DISCUSSION

A. *The relationship between WV2-derived RIs and N*

Table 1 shows the descriptive statistics for the measured N concentrations, for each category of species, as well as the whole-dataset (n = 214). The mean N concentration (1.54%) was largest for *R. altera*, whilst the mean N concentration (0.73%) was lowest for the C3 grass (*F. costata*), which can be considered as a weed [33].

TABLE I
N CONCENTRATIONS (%) IN C3 AND C4 GRASS SPECIES SAMPLED IN THE CATHEDRAL PEAK STUDY AREA (100% DM)

Dependent variables	No. of samples	Mean	Min	Max	SD	CV (%)
Nitrogen (%)						
F. costata	64	0.73	0.41	1.56	0.25	34.42
T. triandra	101	1.11	0.54	1.89	0.37	33.11
R. altera	49	1.54	0.53	1.86	0.24	15.26
All combined	214	1.09	0.41	1.89	0.43	38.87

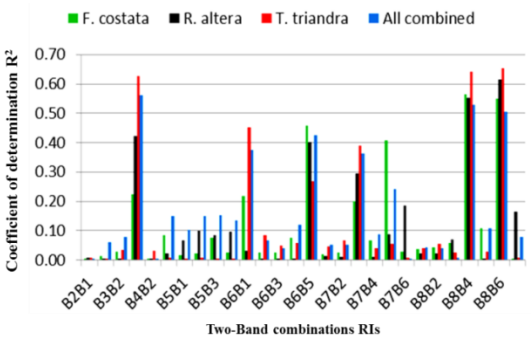

Fig. 2. Coefficient of determination (R^2) between N concentration and RIs. The R^2 values obtained were symmetrical for the all possible two band combinations using the eight multispectral bands of WorldView-2.

Although correlation coefficients between vegetation indices and N vary, similar patterns can be observed between different species. Fig. 2 shows the correlation coefficients between WV2-derieved vegetation indices and N concentration for each species, and across all species. The RI computed involving the costal-blue and yellow bands yielded the highest coefficient of correlation (Fig 2). This can be attributed to strong light absorption due to leaf concentration of chlorophylls *a* and *b* [15]. On the other hand, the maximum sensitivity of chlorophyll *a* concentration is around 520 nm to 630 nm, a range within which the yellow band (585–625 nm) of the WV2 is found. In addition, the RIs computed involving the longer near-infrared bands yielded strong correlations with nitrogen. The near-infrared region has been reported to show strong correlation with concentration of macro-nutrient content in green plants, even at low levels of concentration [50]. The current study also showed that the longer wavelength, near-infrared (860–1040 nm) band combined well with the yellow and red-edge wavebands of WV2 to yield good results. This can be explained by possible influence of inter-species variation [33, 47].

Table 2 shows the results obtained from the RF regression analysis: R^2, RMSE, and NSE values. An analysis of the results indicates that canopy concentration of N is better estimated across all species, compared to within class estimation. This is explained by the relatively low variation in N concentration, in each of the grasslands under study, on one hand, and the magnitude of RF capability for detecting canopy N concentration, on the other hand. This is shown by very low within class NSE (0.42 and 0.18) values obtained for *F. costata* and *R. altera*, respectively. Fig. 3 shows the predicted versus measured N concentration for randomly selected training and test data sets for each target grass species.

TABLE II
PERCENTAGE VARIATIONS ACCOUNTED FOR BY RF REGRESSION MODELS, USING N = 28 PREDICTOR VARIABLES. ALL SPP. (ALL SPECIES).

Grass species	Training dataset (n = 150 for all-combined)			Test dataset (n = 64)		
	R^2	RMSE (%)	NSE	RMSE (%)	NSE	R^2
F. costata	0.45	0.14	0.42	-	-	-
T. triandra	0.56	0.26	0.50	-	-	-
R. altera	0.51	0.20	0.18	-	-	-
All Spp.	0.70	0.25	0.68	0.26	0.68	0.68
All Spp.*	0.71	0.24	0.70	0.24	0.69	0.71

*Selected RIs (n = 6) of two-band combinations: B4B1, B6B1, B8B4, B6B5, B7B3, B8B6, with model mtry = 3 and ntree = 500.

B. *Random forest-based VIP and variable selection*

The RF model implementation inherently determines VIP for each input RI feature. Fig. 4 shows the VIP scores obtained in the current experiment. The order of importance is similar to that of a previous study, which predicted N concentration in the African rangeland environment, using WV2-derived features [20]. RF-based forward variable selection was used to select an optimum subset (n = 6) of RI variables and to assess whether the prediction accuracy could be improved [19, 42]. The result obtained indicates a marginal improvement. This can be explained on the basis

that the RF is insensitive to the relatively small size (n < 100) of the input RI features (n = 28). The use of the RF demonstrated a robust approach of integrating remote sensing applications with field data to predict or map variability in N concentration, across the C3 and C4 type of grasslands.

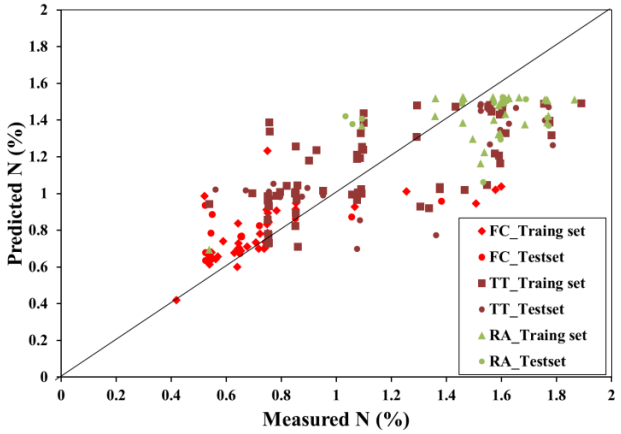

Fig. 3. RF regression using the training and test datasets: *F. costata* (FC), *T. triandra* (TT), *R. altera* (RA).

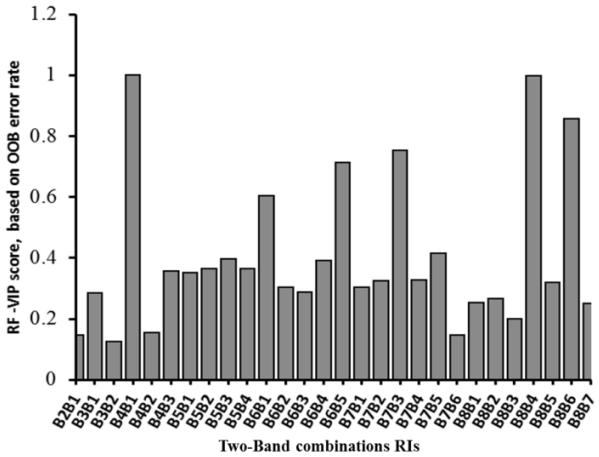

Fig. 4. RF-based VIP for RIs in predicating N. The RF model was developed using mtry = 9 and ntree = 500. Higher score indicates greater OOB error rate when the variable is permuted and thus higher importance.

C. Mapping nitrogen concentration

In in this study the RF regression algorithm was also implemented to map the percentage N concentration across the landscape, composed of the C3 and C4 type of grasses under investigation. The model implementation was done using the EnMAP-box software. Fig. 5 depicts the spatial distribution of the canopy N concentration in the study area.

Although the area mapped is underlain by relatively homogeneous Drakensberg basalt formations of the Stormberg series, there are, undoubtedly, spatial variations in canopy N concentrations, as depicted in the final map. Previous studies conducted in the area have shown that there are strong associations of *T. triandra* and *F. co*stata occurrence on warm northerly and cool southerly slopes, respectively [32]. Strikingly, the final map of N tends to reflect the mean nutrient concentration between the *T. triandra* and *F. costata*, as observed in the study area. Moreover, the good NSE (Table 2) value obtained for the All-combined data indicates that there exists the capability of detecting N concentration in the areas, given the inter-species variation. In addition, the spatial distribution of the N values, as predicted for the high-lying crests, which were dominated by the T. *triandra* and *R. altera*, also showed higher concentrations of N. The final map thus offers the potential for linking forage N variability with range conditions, even in landscapes with disparate C3 and C4 grass/weed composition.

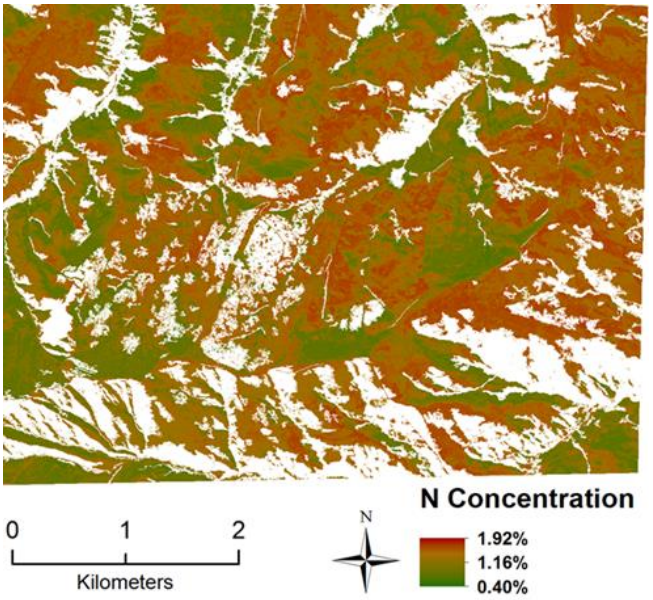

Fig. 5. EnMAP-Box implementation of RF for the estimation of N distribution in the study area (map is 2 m spatial resolution). Light green pixels represent the lowest N concentration values. The white spaces represent masked or other land-cover types that were not modeled. Further details on the EnMAP-Box software utility can be found on the website of the EnMAP: http://www.enmap.org.

V. CONCLUSION

The main focus was to predict and map C3 and C4 grass canopy concentration of N using vegetation indices based on a combination of bands developed from new generation remote sensing multispectral data. Random forest analyses of the canopy N concentration using WV2 imagery yielded statistically significant prediction model, which produced an accurate map of canopy N for the Cathedral Peak study area in South Africa. The prediction of canopy N concentration across C3 and C4 grasslands in the area is within 0.22% of the mean observed N content. The result showed that the study approach is robust and generally applicable where canopy N data are available for C3 and C4 grass environment. In addition, this study demonstrates the

practical utility of WV2 multispectral satellite imaging sensor and for future strategically positioned, few bands, imaging sensors for mapping biochemical properties of vegetation. This potential means that scientists and land managers will be able to make spatial estimates of N variability that is critical to understanding the functioning of ecosystems. Nevertheless, the effects of other canopy properties such as biomass and phenology on canopy N estimation remain a subject of further investigation.

REFERENCES

[1] J. C. von Fischer, L. L. Tieszen, and D. S. Schimel, "Climate controls on C_3 vs. C_4 productivity in North American grasslands from carbon isotope composition of soil organic matter," *Global Change Biology*, vol. 14, pp. 1141-1155, 2008.

[2] S. J. E. Wand, G. F. Midgley, M. H. Jones, and P. S. Curtis, "Responses of wild C4 and C3 grass (Poaceae) species to elevated atmospheric CO_2 concentration: a meta-analytic test of current theories and perceptions," *Global Change Biology*, vol. 5, pp. 723-741, 1999.

[3] J. C. Winslow, E. R. Hunt Jr, and S. C. Piper, "The influence of seasonal water availability on global C3 versus C4 grassland biomass and its implications for climate change research," *Ecological Modelling*, vol. 163, pp. 153-173, 2003.

[4] R. V. Barbehenn, Z. Chen, D. N. Karowe, and A. Spickard, "C3 grasses have higher nutritional quality than C_4 grasses under ambient and elevated atmospheric CO_2," *Global Change Biology*, vol. 10, pp. 1565-1575, 2004.

[5] D. Chen, H. W. Hunt, and J. A. Morgan, "Responses of a C3 and C4 perennial grass to CO_2 enrichment and climate change: Comparison between model predictions and experimental data," *Ecological Modelling*, vol. 87, pp. 11-27, 1996.

[6] S. Niu, W. Liu, and S. Wan, "Different growth responses of C3 and C4 grasses to seasonal water and nitrogen regimes and competition in a pot experiment," *Journal of Experimental Botany*, vol. 59, pp. 1431-1439, April 1, 2008 2008.

[7] P. B. Reich, D. Tilman, J. Craine, D. Ellsworth, M. G. Tjoelker, J. Knops, D. Wedin, S. Naeem, D. Bahauddin, J. Goth, W. Bengtson, and T. D. Lee, "Do species and functional groups differ in acquisition and use of C, N and water under varying atmospheric CO_2 and N availability regimes? A field test with 16 grassland species," *New Phytologist*, vol. 150, pp. 435-448, 2001.

[8] R. F. Sage and R. W. Pearcy, "The Nitrogen Use Efficiency of C3 and C4 Plants: II. Leaf Nitrogen Effects on the Gas Exchange Characteristics of Chenopodium album (L.) and Amaranthus retroflexus (L.)," *Plant Physiology*, vol. 84, pp. 959-963, 1987.

[9] O. Beeri, R. Phillips, J. Hendrickson, A. B. Frank, and S. Kronberg, "Estimating forage quantity and quality using aerial hyperspectral imagery for northern mixed-grass prairie," *Remote Sensing of Environment*, vol. 110, pp. 216-225, 2007.

[10] W. Bond, J., C. J. Geldenhuys, T. M. Everson, C. S. Everson, and M. F. Calvin, "Fire ecology: characteristics of some important biomes of sub-Sahara Africa," in *Wildland Fire Management Handbook for Sub-Sahara Africa*, J. G. Goldammer and C. de Ronde, Eds., ed: Global Fire Monitoring Center, 2004, pp. 11-26.

[11] J. Grant, C. Hopcraft, H. Olff, and A. Sinclair, "Herbivores, resources and risks: Alternating regulation along primary environmental gradients in savannas.," *Trends in Ecology & Evolution*, vol. 25, p. 119−128, 2010.

[12] N. M. Knox, A. K. Skidmore, G. P. Asner, H. H. T. Prins, H. M. A. van der Werff, W. F. de Boer, C. van der Waal, H. J. de Knegt, E. M. Kohi, R. Slotow, and C. C. Grant, "Mapping savanna forage quality, in the dry season, using CAO Alpha imagery," *Remote Sensing of Environment*, vol. 115, pp. 1478-1488, 2011.

[13] C. Adjorlolo, O. Mutanga, M. A. Cho, and R. Ismail, "Challenges and opportunities in the use of remote sensing for C3 and C4 grass species discrimination and mapping," *African Journal of Range & Forage Science* vol. 29, pp. 47-61, 2012.

[14] S. L. Ustin and J. A. Gamon, "Remote sensing of plant functional types," *New Phytologist*, vol. 186, pp. 795-816, 2010.

[15] P. M. Hansen and J. K. Schjoerring, "Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression," *Remote Sensing of Environment*, vol. 86, pp. 542-553, 2003.

[16] S. Ullah, Y. Si, M. Schlerf, A. K. Skidmore, M. Shafique, and I. A. Iqbal, "Estimation of grassland biomass and nitrogen using MERIS data," *International Journal of Applied Earth Observation and Geoinformation*, vol. 19, pp. 196-204, 2012.

[17] L. Kumar, K. S. Schmidt, S. Dury, and A. K. Skidmore, "Review of hyperspectral remote sensing and vegetation science.," in *Hyperspectral remote sensing*, F. Van Der Meer, Ed., ed Dordrecht: Kluwer Academic Press, 2001.

[18] M. A. Cho, R. Mathieu, G. P. Asner, L. Naidoo, J. van Aardt, A. Ramoelo, P. Debba, K. Wessels, R. Main, I. P. J. Smit, and B. Erasmus, "Mapping tree species composition in South African savannas using an integrated airborne spectral and LiDAR system," *Remote Sensing of Environment*, vol. 125, pp. 214-226, 2012.

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

[20] F. M. Zengeya, O. Mutanga, and A. Murwira, "Linking remotely sensed forage quality estimates from WorldView-2 multispectral data with cattle distribution in a savanna landscape," *International Journal of Applied Earth Observation and Geoinformation* vol. 21, pp. 513-524, 2013.

[21] M. A. Cho and A. K. Skidmore, "Hyperspectral predictors for monitoring biomass production in Mediterranean mountain grasslands: Majella National Park, Italy," *International Journal of Remote Sensing*, vol. 30, pp. 499-515, 2009.

[22] F. Fava, R. Colombo, S. Bocchi, M. Meroni, M. Sitzia, N. Fois, and C. Zucca, "Identification of hyperspectral vegetation indices for Mediterranean pasture characterization," *International Journal of Applied Earth Observation and Geoinformation*, vol. 11, pp. 233-243, 2009.

[23] J. G. Ferwerda, A. K. Skidmore, and O. Mutanga, "Nitrogen detection with hyperspectral normalized ratio indices across multiple plant species," *International Journal of Remote Sensing*, vol. 26, pp. 4083-4095, 2005.

[24] D. Zhao, K. R. Reddy, V. G. Kakani, and V. R. Reddy, "Nitrogen deficiency effects on plant growth, leaf photosynthesis, and hyperspectral reflectance properties of sorghum," *European Journal of Agronomy*, vol. 22, pp. 391-403, 5// 2005.

[25] D. L. Peterson and G. S. Hubbard, "Scientific Issues and potential remote-sensing requirements for plant biochemical content," *Journal of Imaging Science and Technology*, vol. 36, pp. 446-456, 1992.

[26] T. Kavzoglu and P. M. Mather, "The role of feature selection in artificial neural network applications," *International Journal of Remote Sensing*, vol. 23, pp. 2919-2937, 2002/01/01 2002.

[27] J. G. P. W. Clevers, G. W. A. M. van der Heijden, S. Verzakov, and M. E. Schaepman, "Estimating grassland biomass using SVM band shaving of hyperspectral data," *Photogrammetric Engineering & Remote Sensing*, vol. 73, pp. 1141-1148, 2007.

[28] M. Dalponte, L. Bruzzone, L. Vescovo, and D. Gianelle, "The role of spectral resolution and classifier complexity in the analysis of hyperspectral images of forest areas," *Remote Sensing of Environment*, vol. 113, pp. 2345-2355, 2009.

[29] J. C.-W. Chan and D. Paelinckx, "Evaluation of random forest and adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral

imagery," *Remote Sensing of Environment,* vol. 112, pp. 2999-3011, 2008.

[30] D. J. B. Killick, *An account of the plant ecology of the Cathedral Peak area of the Natal Drakensberg* vol. 34, 1963.

[31] J. E. Granger and R. E. Schulze, "Incoming Solar Radiation Patterns and Vegetation Response: Examples from the Natal Drakensberg," *Vegetatio,* vol. 35, pp. 47-54, 1977.

[32] T. R. Hill, "Description, classification and ordination of the dominant vegetation communinities, Cathedral Peak, KwaZulu-Natal Drakensberg," *South African Journal of Botany,* vol. 62, pp. 263-269, 1996.

[33] C. Adjorlolo, O. Mutanga, A. M. Cho, and R. Ismail, "Spectral resampling based on user-defined inter-band correlation filter: C3 and C4 grass species classification," *International Journal of Applied Earth Observation and Geoinformation,* vol. 21, pp. 535-544, 2013.

[34] C. Adjorlolo and O. Mutanga, "Integrating remote sensing and geostatistics to estimate woody vegetation in an African savanna," *Journal of Spatial Science,* pp. 1-18, 2013.

[35] C. Adjorlolo, O. Mutanga, and A. M. Cho, "Mapping and characterizing C3 and C4 grass distribution in the subtropical montane rangeland of Southern Africa," *in preparation*.

[36] I. Novozamsky, V. J. G. Houba, R. van Eck, and W. van Vark, "A novel digestion technique for multi-element plant analysis," *Communications in Soil Science and Plant Analysis,* vol. 14, pp. 239-248, 1983/06/01 1983.

[37] Y.-C. Tian, K.-J. Gu, X. Chu, X. Yao, W.-X. Cao, and Y. Zhu, "Comparison of different hyperspectral vegetation indices for canopy leaf nitrogen concentration estimation in rice," *Plant and Soil,* vol. 376, pp. 193-209, 2014/03/01 2014.

[38] T. Fourty, F. Baret, S. Jacquemoud, G. Schmuck, and J. Verdebout, "Leaf optical properties with explicit description of its biochemical composition: Direct and inverse problems," *Remote Sensing of Environment,* vol. 56, pp. 104-117, 5// 1996.

[39] P. J. Curran, "Remote sensing of foliar chemistry," *Remote Sensing of Environment,* vol. 30, pp. 271-278, 1989.

[40] L. Serrano, J. Peñuelas, and S. L. Ustin, "Remote sensing of nitrogen and lignin in Mediterranean vegetation from AVIRIS data: Decomposing biochemical from structural signals," *Remote Sensing of Environment,* vol. 81, pp. 355-364, 2002.

[41] L. Breiman, "Random Forests," *Machine Learning,* vol. 45, pp. 5 - 32, 2001.

[42] R. Ismail and O. Mutanga, "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,* vol. 12, pp. S45-S51, 2010.

[43] U. Grömping, "Variable Importance Assessment in Regression: Linear Regression versus Random Forest," *The American Statistician,* vol. 63, pp. 308-319, 2009/11/01 2009.

[44] A. Hapfelmeier and K. Ulm, "A new variable selection approach using Random Forests," *Computational Statistics & Data Analysis,* vol. 60, pp. 50-69, 4// 2013.

[45] R. Genuer, J.-M. Poggi, and C. Tuleau-Malot, "Variable selection using random forests," *Pattern Recognition Letters,* vol. 31, pp. 2225-2236, 10/15/ 2010.

[46] E. M. Abdel-Rahman, F. B. Ahmed, and R. Ismail, "Random forest regression and spectral band selection for estimating sugarcane leaf nitrogen concentration using EO-1 Hyperion hyperspectral data," *International Journal of Remote Sensing,* vol. 34, pp. 712-728, 2013/01/20 2009.

[47] C. Adjorlolo, M. A. Cho, O. Mutanga, and R. Ismail, "Optimizing spectral resolutions for the classification of C3 and C4 grass species, using wavelengths of known absorption features," *Journal of Applied Remote Sensing,* vol. 6, 2012.

[48] J. G. N. Irisarri, M. Oesterheld, S. R. Verón, and J. M. Paruelo, "Grass species differentiation through canopy hyperspectral reflectance," *International Journal of Remote Sensing,* vol. 30, pp. 5959 - 5975, 2009.

[49] R. R. Pullanagari, I. J. Yule, M. P. Tuohy, M. J. Hedley, R. A. Dynes, and W. M. King, "In-field hyperspectral proximal sensing for estimating quality parameters of mixed pasture " *Precision Agriculture,* vol. 13, pp. 351-369, 2012.

[50] M. A. Cho, A. K. Skidmore, and C. Atzberger, "Towards red-edge positions less sensitive to canopy biophysical parameters for leaf chlorophyll estimation using properties optique spectrales des feuilles (PROSPECT) and scattering by arbitrarily inclined leaves (SAILH) simulated data," *International Journal of Remote Sensing,* vol. 29, pp. 2241-2255, 2008/04/01 2008.

Clement Adjorlolo received the undergraduate qualification in Wildlife Management from the College of African Wildlife Management, Tanzania, in 2001, and received the B.S. Honours, MSc. and PhD degree in Physical Geography and Environmental Science from the University of KwaZulu-Natal, South Africa, in 2007, 2009 and 2013, respectively. His research focus in recent years has been on the application of remote sensing and Geographic Information System (GIS) technology to rangeland resources assessment and monitoring. He is currently a Senior Scientist with the South African National Space Agency (SANSA), involved in the development of spatial statistical, multispectral and hyperspectral techniques for modelling biophysical and biochemical properties of rangeland vegetation and for mapping vegetation associations.

Onisimo Mutanga is a full Professor and Academic Leader (Research) in the School of Agriculture, Earth and Environmental Science at the University of KwaZulu- Natal. He completed his PhD on Hyperspectral Remote Sensing of Tropical grass quality and quantity at Wageningen University-ITC (Netherlands) in 2004. His expertise lies on Ecological Remote Sensing, with particular emphasis on vegetation pattern analysis and monitoring, agricultural landuse mapping as well as wildlife habitat evaluation. His focus in recent years has been on the development of remote sensing techniques for mapping tropical vegetation quality and quantity to understand wildlife feeding patterns and distribution. His emerging niche areas include mapping vegetation species, disease infestation on plantation forests and agricultural crops as well as quantification of forest degradation and its impact on biodiversity and ecosystem condition. Prof Mutanga has graduated 7 PhDs and 14 Masters Students and is currently supervising more than 15 post graduate students. He has published more than 50 articles in ISI journals and has several conference proceedings and book chapters. He holds a C1 NRF rating and serves in several national and international research committees.

Moses A. Cho received the B.Sc. degree in Natural sciences (option: botany) from the University of Yaoundé, Cameroon in 1991. He received the M.Sc. degree in biodiversity conservation in 2001 from the Greenwich University, UK. He obtained his Ph.D. degree in 2007 from the Wageningen University and The International Institute for Geoinformation Science and Earth Observation (ITC), The Netherlands. He is currently a principal research scientist with the Earth Observation (EO) group at The Council for Scientific and Industrial Research (CSIR), South Africa and a research fellow with the University of KwaZulu-Natal South Africa. His current research interest involves the use of hyperspectral and new specialised spaceborne multispectral imagery for assessing floral diversity, and vegetation productivity at a regional scale.