Preprint

Very high resolution aboveground carbon mapping in subtropical thicket

Dugal Harris
Cosman Bolus
James Reeler

Published paper:

http://dx.doi.org/10.1117/1.JRS.15.038502

Source code:

https://github.com/dugalh/map_thicket_agc

Cite as:

Dugal Harris, Cosman Bolus, James Reeler, "Very high resolution aboveground carbon mapping in subtropical thicket," *Journal of Applied Remote Sensing.* **15**(3), 038502 (2021), doi: 10.1117/1.JRS.15.038502

Very high resolution aboveground carbon mapping in subtropical thicket

Dugal Harris,[a*] Cosman Bolus,[b] James Reeler,[c]
[a]Stellenbosch University, Department of Geography and Environmental Studies, Stellenbosch, South Africa, 7602
[b]Sustainable Landscape Services, Muizenberg, Cape Town, South Africa, 7945
[c]WWF South Africa, Newlands, Cape Town, South Africa, 7700

Abstract. Aboveground carbon (AGC) maps will assist with the monitoring and verification of carbon stored through restoration of subtropical thicket. Field methods are required for capturing AGC ground truth data at the plot scale, but are highly impractical for large area mapping of carbon stocks. Against this background, a remote sensing method to estimate AGC from very high spatial resolution multi-spectral imagery was developed. AGC ground truth was acquired for 85 plots in a small study area in the Baviaanskloof, South Africa. Using the ground truth, univariate and multivariate snapshot models were developed to predict AGC from features in a 2017 WorldView-3 (0.34 m resolution 8-band) satellite image. Informative features were selected from a large set of spectral, textural and vegetation index features using stepwise forward selection. Using this approach, the multivariate model produced a coefficient of determination (R^2) of 0.886 and root mean square error of 2.862 t C ha^{-1}. This study demonstrates the efficacy of regression approaches for estimating AGC from multi-spectral imagery, and provides a foundation for the spatial and temporal extension of AGC remote sensing in the thicket biome.

Keywords: biomass, aboveground carbon, subtropical thicket, worldview-3, very high resolution, regression

***Dugal Harris, E-mail: dugalh@gmail.com

1 Introduction

1.1 Restoration of subtropical thicket

The subtropical thicket biome occurs in the southern and south-eastern parts of South Africa[1] and is characterized by a dense and diverse combination of small trees, thorny shrubs, grasses and succulents.[2] Degradation of this semi-arid biome, caused predominantly by poorly managed goat browsing, is associated with severe reductions in biodiversity and ecological functioning.[3–5] It is estimated that 90% of the thicket biome has suffered moderate to severe degradation of this form.[5,6] Degraded habitats do not recover after removal of livestock,[7] and active measures are required to restore ecosystem health.

Studies at established restoration sites have suggested that planting of spekboom (*Portulacaria afra*), a dominant and keystone species,[8,9] is a viable means for large-scale thicket restoration.[10] Spekboom is unusually effective at sequestering carbon for an arid region plant[10], which has obvious significance for climate change mitigation. Approximately 1.7 million ha of thicket is suitable for restoration in this manner.[5]

Spatial information is required to support the planning and monitoring of restoration. In the context of climate change, maps of above ground carbon (AGC), an important indicator of ecosystem health, are of particular interest.[11–13] Carbon stored through restoration can be verified and traded on the carbon market, providing a source of restoration funding.[14,15] There is also the possibility of thicket restoration qualifying for offsets under South Africa's carbon tax.[16,17] Field methods for quantifying carbon stocks are time-consuming and costly, and highly impractical for large scale carbon stock mapping.[18,19] This is especially true in the thicket biome due to rugged terrain, dense vegetation and complex growth forms.[13,20] A remote sensing approach for AGC mapping in thicket would therefore be of great assistance in the cost-effective monitoring and auditing of carbon stocks generated through restoration.

1.2 Remote sensing of biomass

Various forms of light detection and ranging (LiDAR), synthetic aperture radar (SAR) and optical (multi-spectral and hyperspectral) data have been used (independently or in combination) for remote sensing of AGC and biomass.[21] LiDAR data provides information on the three-dimensional spatial structure of vegetation.[18] Given the known correlation between vegetation spatial dimension and biomass,[20] LiDAR has obvious application for biomass estimation. Very high spatial resolution (VHSR) LiDAR is however costly and limited in coverage.[22] Its suitability for use in short-stature vegetation is hampered by difficulties in separating canopy from ground

returns, and relatively low penetration of dense canopies,[23] such as those of intact subtropical thicket.

SAR data is often used for biomass mapping, especially in forested areas.[21] A variety of radar wavelengths are used for SAR, each being sensitive to different components of vegetation structure.[22] Long (L- and P-band) wavelengths are the most frequently used for measuring biomass.[18,21,24] Drawbacks of SAR include sensitivity to soil moisture and topography,[18,24] spatial resolution limitations in the L- and P-bands,[18] and the need for complex pre-processing.[22] At present the effectiveness of SAR for mapping thicket biomass is untested.

Optical reflectance in the visible, near-infrared (NIR) and short-wave infrared (SWIR) bands is affected by vegetation structure and photosynthetic activity, and is indirectly correlated with biomass.[18,21] Optical satellite data is cost-effective, has global coverage, and is increasingly available in a variety of spectral and spatial resolutions. Drawbacks of optical data are its dependence on atmospheric conditions and viewing geometry (anisotropy),[25] and its lack of sensitivity to the vertical dimension i.e. vegetation height. Saturation of optical measurements can occur in high biomass forest,[21] but this is rare in semi-arid habitats,[18] such as thicket.

In the majority of cases, carbon stocks are estimated using regression models that describe empirical relationships between features (i.e. variables) derived from remote sensing data and carbon stock ground truth.[21,22] In some LiDAR studies, a more theoretical approach is used where variables such as canopy height and area can be measured directly.[23] Regression methods are divided into parametric and non-parametric types. Parametric approaches assume a model structure *a priori* and use the data to estimate the parameters of this model (e.g. linear regression). In non-parametric approaches, the model structure is effectively learnt from the data and no assumptions about the form of the underlying relationships are made.[26] The non-parametric

approach is suited to cases where the underlying model is complex and non-linear, as these relationships are seldom known in advance. However, as a general rule, the non-parametric approach requires more data for representative fitting,[21] and the parametric approach is better suited to limited ground truth data sets.

1.3 Remote sensing of subtropical thicket

Thompson et al.[6] produced a degradation map of the biomes occurring in the Little Karoo. A straightforward thresholding approach, using 250 m resolution MODIS normalized difference vegetation index (NDVI) data, was used for mapping degradation in thicket. The study was successful at estimating three degradation levels (intact, moderate and severe) of spekboom thicket.

A method for estimating AGC in subtropical thicket was devised by Nyamugama & Kakembo[27] using 10 m SPOT 5 imagery and a set of 90 30 m × 30 m carbon stock ground truth plots. A univariate quadratic regression model was fitted using a SPOT-5 NDVI feature (dependent variable) which achieved R^2 values of 0.975, 0.812 and 0.725 for intact, transformed and degraded classes respectively. The SPOT-5 model was subsequently applied unmodified to a time series of Landsat NDVI, to estimate change in carbon stocks over the period 1972-2010. Time series results were not validated, but showed a marked decrease in AGC over time.

A method for regional mapping of spekboom canopy cover was developed by Harris, Vlok & Van Niekerk[28]. Multi-spectral VHSR aerial imagery from National Geo-spatial Information (NGI)[29] of a large region in the Little Karoo was corrected for spatially varying atmospheric and anisotropic effects using a radiometric homogenization procedure.[30] Feature clustering and ranking (FCR)[31] was applied to select an informative set of features from a high dimensional set

of spectral, vegetation index and texture features. A supervised classification (decision tree) approach was subsequently applied to the reduced feature set, to produce the canopy cover map.

In general, most biomass mapping has been focused on forests, and relatively few studies have addressed arid habitats,[18,21] and thicket in particular. The biodiversity of thicket, and rugged terrain in which it occurs,[32] contribute to unwanted variation in remotely sensed images,[30,33] and make vegetation mapping in this biome a challenge.

1.4 Contribution

In this paper we present the development and evaluation of a technique for the remote sensing of AGC in subtropical thicket. Practical constraints of budget, availability, and spatial resolution, meant that SAR and LiDAR were not considered in this study. WorldView-3 8-band multi-spectral imagery of a small (3000 ha) study site in the Baviaanskloof was the principal source of data. Multi-spectral imagery has been used successfully in related studies.[27,28] WorldView-3 imagery provides four bands in the red-infrared spectral region which is known to be informative for vegetation characterization.[34] The high WorldView-3 spatial resolution of 0.34 m facilitated the use of features describing vegetation texture and allowed detailed coverage of small (10 m × 10 m) ground truth plots. A set of vegetation index, band ratio and texture features (and non-linear transformations thereof) were generated from the imagery and reduced to an informative minimum using a stepwise forward selection approach.[26] Linear regression models were subsequently fitted to selected features and evaluated with AGC ground truth data comprising 85 plots.

2 Data

2.1 Study site

The study area is located in the central Baviaanskloof, South Africa (see Fig. 1). Subtropical thicket, fynbos and renosterveld vegetation types are all found in the study area.[35] Spekboom thicket on the mid to lower slopes in this area has been severely degraded by over-browsing. The less accessible, and more rugged upper slopes in the western section contain areas of intact and moderately degraded thicket.

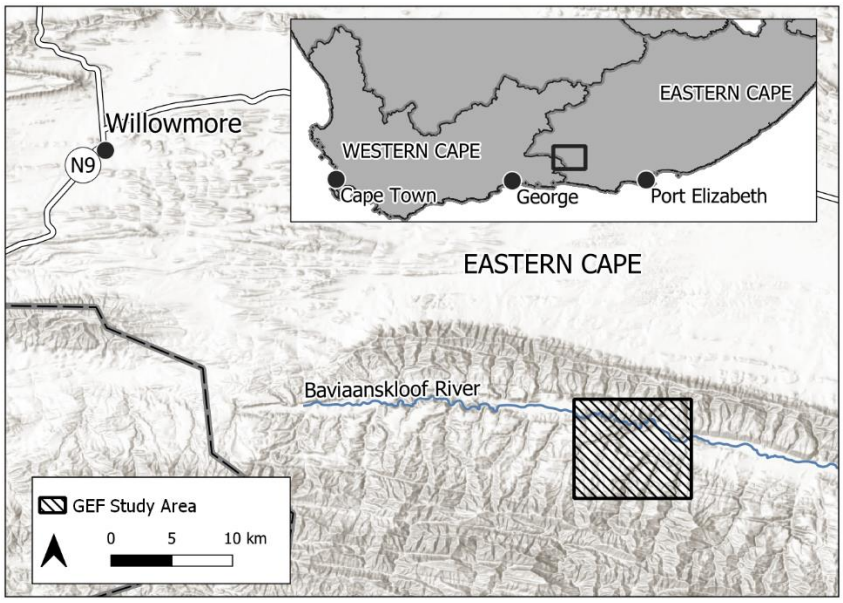

Fig. 1 Study area context map

2.2 Imagery

Two images of the study area were obtained, consisting of a WorldView-3 0.34 m 8-band satellite image, and a 0.5 m 4-band stereo aerial mosaic from NGI.[29] Details of the imagery are shown in Table 1. A set of 30 ground control points (GCPs) of clearly distinguishable ground features were gathered in situ with a differential global positioning system (DGPS) device and post-processed to ±20 cm accuracy. Using the GCPs in combination with the NGI 2015 stereo imagery, a VHSR

(0.5 m) digital surface model (DSM) was produced with the Agisoft Photoscan[36] photogrammetry application. The GCP locations and DSM are shown in Fig. 3. Using the GCPs and DSM, the WV3 2017 image was orthorectified to sub-meter accuracy, and subsequently corrected for atmospheric effects using ATCOR-3[37]. The multispectral bands were pan-sharpened to 0.34 m resolution using the PCI-Geomatica[38] PANSHARP2 algorithm, with cubic convolution resampling. This algorithm has been shown to provide similar or superior spectral and spatial reconstruction compared to other methods.[39,40]

Table 1 Imagery details

Abbreviation	Source	Spatial resolution	Spectral bands	Component images	Date	Image nadir	Sun elevation
WV3 2017	WorldView-3	0.34 m	8	1	2017/10/01	16.9°	53.7°
NGI 2015	NGI aerial	0.50 m	4	4	2015/04/27	-5.0° – 5.0°	>30.0°

2.3 AGC ground truth

Plant allometric measurements and litter samples were gathered in a series of field sampling exercises between November 2017 and March 2019.[41] Some changes in AGC may have occurred between the date of the WV3 2017 image capture (October 2017) and completion of field sampling. Van der Vyver & Cowling[20] give the mean rate of AGC change in actively restored thicket as 2.3 t C ha^{-1} yr^{-1}. As the field plots were left undisturbed for the duration of the field sampling, any changes in AGC would likely be substantially less than this, and are not considered significant.

A total of 85 sampling plots were spread over intact, moderate and severe thicket degradation strata in a stratified random sampling approach. The severe degradation stratum (defining areas suitable for restoration by spekboom planting) was delineated in situ by Vlok[42]. Intact and moderate degradation strata were subsequently added to this map by visual delineation of the WV3

2017 image. Field photos of typical vegetation in the three degradation strata are shown in Fig. 2**,** and stratum spatial extents are displayed in Fig. 3. Stratified random sampling helped ensure ground truth representivity of the range of carbon stocks present in the landscape. Based on the rule-of-thumb that one needs approximately ten times the number of data points as there are features in the model,[26] the sample size of 85 plots allowed for fitting of a reasonably complex model, and falls in the upper range of similar studies.[43] Sampling plot locations are shown in Fig. 3.

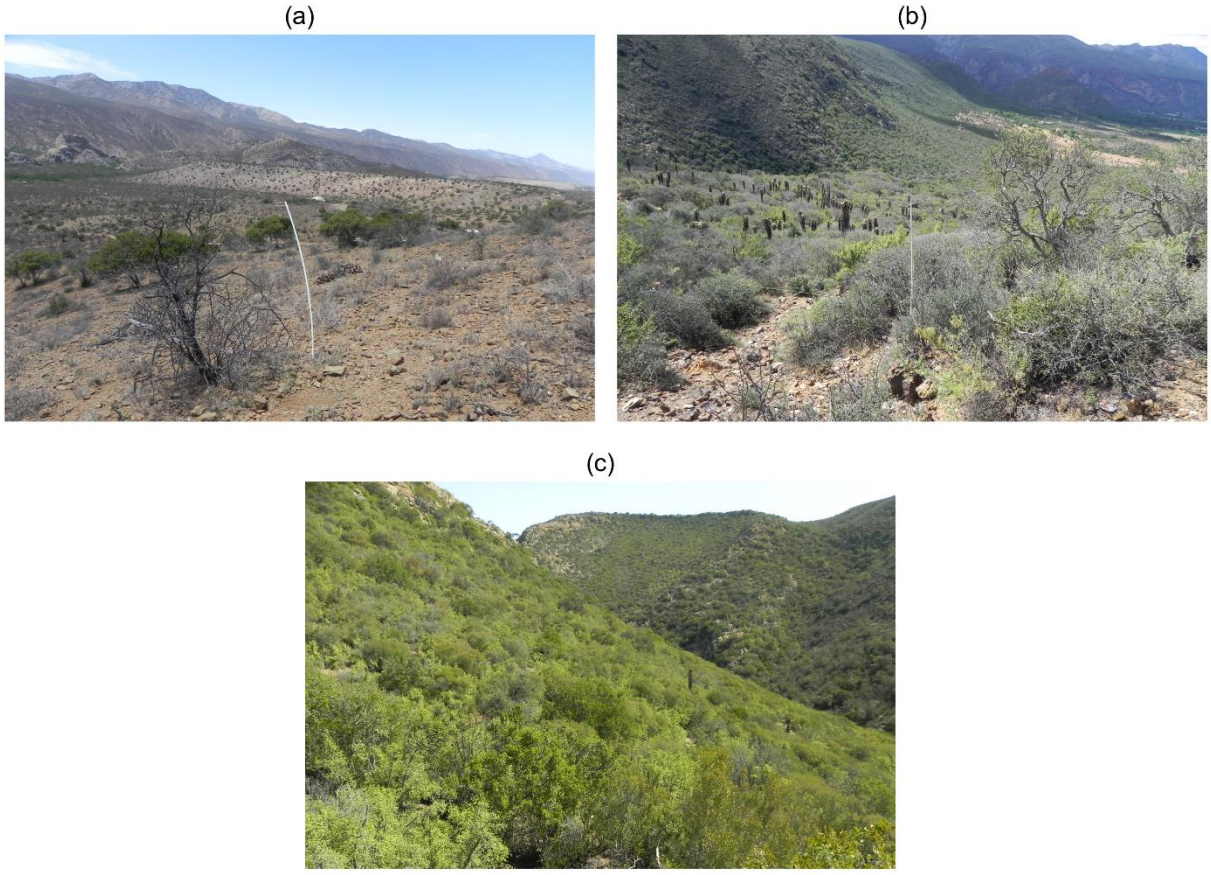

Fig. 2 Typical vegetation in (a) severe, (b) moderate, and (c) intact degradation strata

Plot sizes were chosen as 10 m × 10 m for the moderate and intact strata, and as 20 m × 20 m for the more sparsely vegetated severe stratum, based on previous work.[13] While larger plot sizes

are associated with improved estimation accuracy, reduced variability, and decreased sensitivity to spatial inaccuracy,[44] these benefits need to be traded off against the time and cost of field measurements. A nested plot design was adopted, where all vegetation was measured in a small 5 m × 5 m subplot, and only vegetation over 0.5 m high was measured in the remainder of the containing plot. The nested plot design reduces field work while having minimal impact on AGC estimation accuracy.[45] Plot corners were recorded with a DGPS device and post-processed to ±30 cm accuracy. Emphasis was placed on the spatial accuracy of orthorectified image and plot locations to mitigate the detrimental effects of the modest plot sizes.[44] In addition, the high spatial resolution of the images gave a coverage of ±850 and ±3400 WorldView-3 pixels for the 10 m × 10 m and 20 m × 20 m plots, respectively. Table 2 contains details of area, number of plots and plot size per stratum. Further details of the field data and sampling approach can be found in Bolus et al.[46] and Bolus et al.[41].

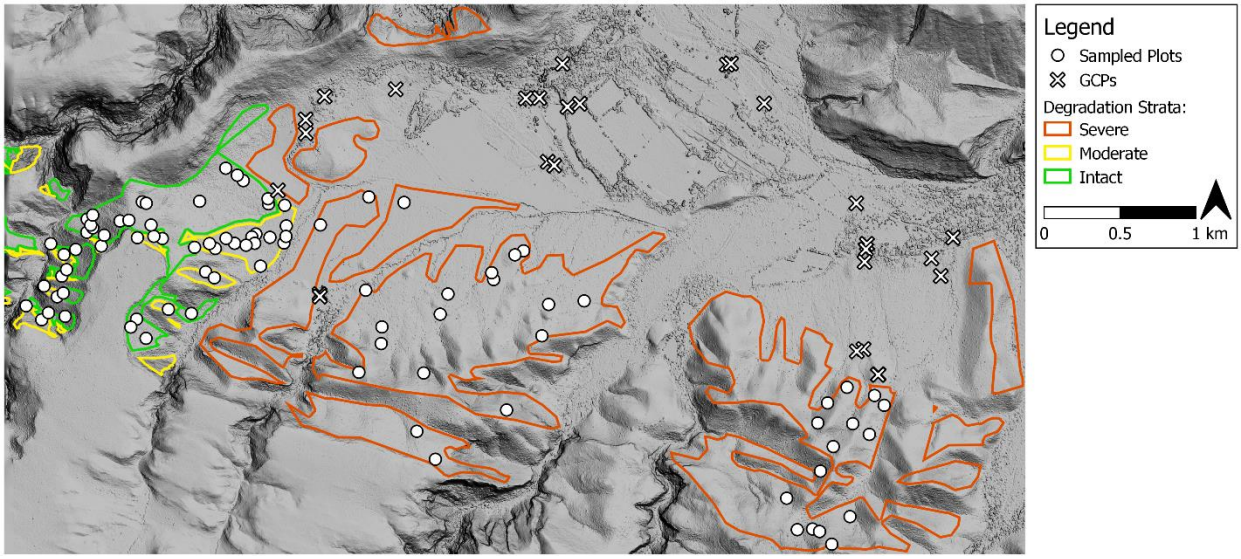

Fig. 3 Study area thicket degradation strata and sampled plots

Table 2 Stratified random sampling details

Degradation stratum	Area (ha)	Number of plots	Nested / containing plot size (m)
Intact	78	26	5×5 / 10×10
Moderate	37	26	5×5 / 10×10
Severe	273	33	5×5 / 20×20
Total	388	**85**	

Aboveground biomass carbon (ABC) was estimated for each plot with species-specific allometric models from Van der Vyver & Cowling[20] and Powell[13]. The allometric models are based on crown diameter and height exclusively, and do not require basal stem diameters which are problematic and time-consuming measurements in thicket.[13,20] Where sampled species had no available allometric data, parameters from a surrogate species of similar genus or growth form were used. Details of the surrogate species map are given in Bolus et al.[41]. ABC for vegetation under 0.5 m high in the containing plot was extrapolated from the 5 m × 5 m nested plot. Litter carbon (C) was estimated by sampling and weighing litter from four 0.5 m × 0.5 m subplots.[41] Aboveground carbon (AGC) was then found as the sum of the ABC and litter C plot components, and extrapolated to per hectare values.

3 Methods

Linear regression models were built and evaluated on features from the WV3 2017 image and the AGC sampling plot data.

3.1 Feature extraction

A large set of features incorporating band ratios, vegetation indices and texture measures were extracted from the WV3 2017 image, which consists of 8 multi-spectral bands and one panchromatic band as detailed in Table 3.

Table 3 WorldView-3 spectral bands

Band number	Band name	Symbol	Wavelength (nm)
1	Panchromatic	pan	450 - 800
2	Coastal	C	400 - 450
3	Blue	B	450 - 510
4	Green	G	510 - 580
5	Yellow	Y	585 - 625
6	Red	R	630 - 690
7	Red-Edge	RE	705 - 745
8	NIR1	$NIR1$	770 - 895
9	NIR2	$NIR2$	860 - 1040

Generated features included ratios of all pair-wise band combinations:

$$BR_{i,j} = \frac{C_i}{C_j} \quad (1)$$

where C_i are the individual bands, i, j are band numbers and $i \neq j$. By including the panchromatic band in the set from which band ratios were generated, intensity normalized bands (e.g. $BR_{4,1} = R/pan$) could be extracted. These normalized features help reduce intensity variation caused by topography and anisotropy.[47]

In the visible spectrum, vegetation reflectance is mainly affected by chlorophyll, which absorbs incident blue and red light while reflecting some green light, giving living vegetation its characteristic green color.[48] Leaves are strongly reflective in the NIR spectrum due to scattering by internal plant structures. The sharp transition from absorption to reflectance around 700 nm is an important feature for characterizing vegetation.[34] The normalized difference vegetation index (NDVI) aims to capture this spectral feature as a single number that is invariant to illumination intensity. The soil adjusted vegetation index (SAVI) modifies NDVI to decrease sensitivity to soil influences.[49] NDVI and SAVI are defined in Equations (2) and (3) respectively. Separate NDVI and SAVI values were calculated using each of the three WorldView-3 NIR bands (i.e. red-edge,

NIR1 and NIR2 as detailed in Table 3). Following Huete[49], a value of 0.5 was used for L in Equation (3).

$$NDVI_i = \frac{NIR_i - R}{NIR_i + R} \tag{2}$$

$$SAVI_i = \frac{(1 + L)(NIR_i - R)}{NIR_i + R + L} \tag{3}$$

where NIR_i = Red-edge, NIR1 or NIR2 for the Worldview-3 image.

Band ratio (*BR*), *NDVI* and *SAVI* per-pixel features were converted to scalar per-plot features using entropy, mean and standard deviation statistics. The entropy statistic describes the amount of randomness in a variable. Entropy and standard deviation serve as basic texture features describing local spectral complexity. The entropy of the vector of sampling plot pixel values, $\boldsymbol{f}$, is defined as[50]

$$Entropy(\mathbf{f}) = -\sum_i h_i(\boldsymbol{f}) \log_2 h_i(\boldsymbol{f}) \tag{4}$$

where $h_i(\boldsymbol{f})$ is the probability in the i^{th} histogram bin of $\boldsymbol{f}$ (100 bins were used).

The mean and standard deviation of the N values in $\boldsymbol{f}$ are defined by Equations (5) and (6) respectively.

$$\bar{\boldsymbol{f}} = \frac{1}{N}\sum_{i=1}^{N} f_i \tag{5}$$

$$Std(\boldsymbol{f}) = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(f_i - \bar{\boldsymbol{f}}\right)^2} \tag{6}$$

In addition to the above features, square, square root and log functions of the mean feature values (from Equation (5)) were included in the extracted set. A total of 522 features were generated with the above procedure.

3.2 Feature selection

For a finite number of samples, increasing the number of model features beyond a certain point results in overfitting and a decrease in regression accuracy. This so-called "peaking phenomenon"[26] requires the size of the feature set to be reduced to a salient minimum in order to achieve accurate regression. Informative subsets of features were identified using a stepwise forward selection (FS) procedure, with the root mean square error (RMSE) of a linear model as evaluation criterion. FS starts with an empty feature set and proceeds in a series of steps where one feature is added to the selected set at each step.[26] The feature whose addition most improves the evaluation criterion is the one added at that step. To avoid positively biasing the RMSE criterion, it was found with five-fold cross-validation.[51] The number of selected features for the multivariate model was chosen to minimize the RMSE criterion.

3.3 Linear regression

Multivariate and univariate linear regression models were fitted to selected features. The univariate model was included as a benchmark, and for use in future studies addressing temporal transferability and spatial extension of thicket AGC maps. A parametric modelling approach was adopted due to the limited number of sampling plots.[22] While the models are linear in the parameters, the inclusion of square root, square and log feature transformations allowed for description of non-linear relationships. The models were evaluated with leave-one-out cross validation (LOOCV) to avoid overfitting. Snapshot AGC maps of the study area were generated by applying the univariate and multivariate models to the WV3 2017 image.

4 Results

Species-specific allometric models of Van der Vyver & Cowling[20] and Powell[13] were applied to field allometric measurements to produce AGC ground truth estimates for each of the 85 plots. These are shown against the WV3 2017 image in a color-infrared rendering, in Fig. 4. Table 4 shows the per-stratum and overall litter C, ABC and AGC mean and standard deviation statistics.

Satellite image courtesy of the DigitalGlobe Foundation

Fig. 4 AGC measurements against WV3 2017 background

Table 4 Per-stratum and total ABC, litter C and AGC statistics

	ABC (t C ha^{-1}) [a]	litter C (t C ha^{-1}) [a]	AGC (t C ha^{-1}) [a]	N [a]
Intact	21.198 (5.972)	9.184 (3.570)	30.381 (7.799)	26
Moderate	13.916 (5.905)	4.033 (2.313)	17.949 (6.770)	26
Severe	6.393 (3.689)	1.939 (2.550)	8.332 (5.207)	33
Total	13.223 (8.021)	4.796 (4.148)	18.018 (11.236)	85

[a] mean (standard deviation), N = number of plots

Fig. 5 shows the change in AGC prediction accuracy (R^2 and negative RMSE) as successive features are added to the model with the stepwise forward selection procedure. The multivariate model was selected to have 15 features where RMSE is at a minimum and R^2 is near its maximum. RMSE was evaluated using LOOCV, and R^2 found from the stacked predictions of the leave-one-out procedure.

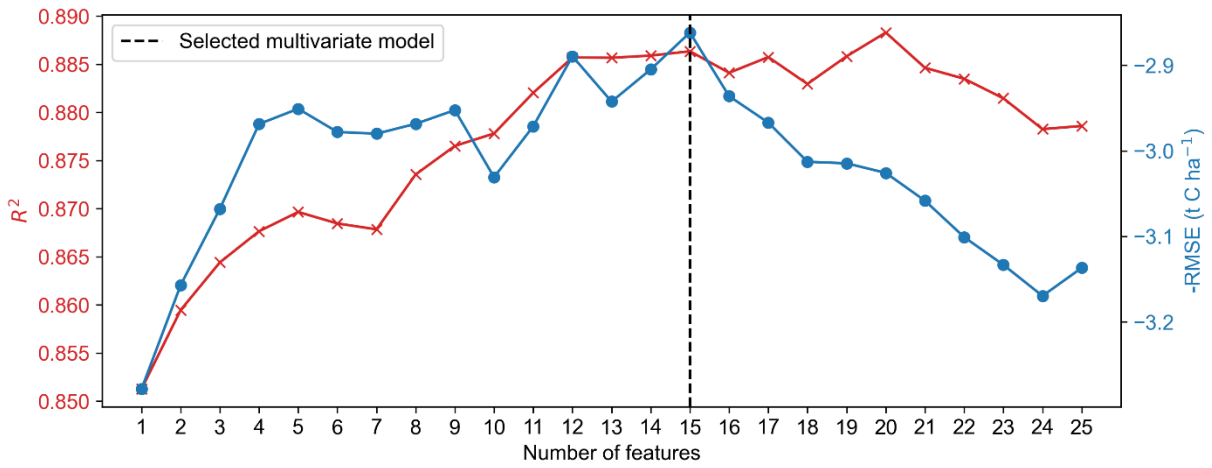

Fig. 5 Change in model accuracy with number of features (R^2 in red and negative RMSE in blue)

The 15 features selected for multivariate model are detailed in Table 5 (band symbols are defined in Table 3). The best univariate model consists of the $Log\left(\overline{R/pan}\right)$ feature (i.e. the first feature selected). Measured AGC is plotted against predicted AGC for the best multivariate and univariate models in Fig. 6. Predicted AGC values were again obtained from the LOOCV procedure.

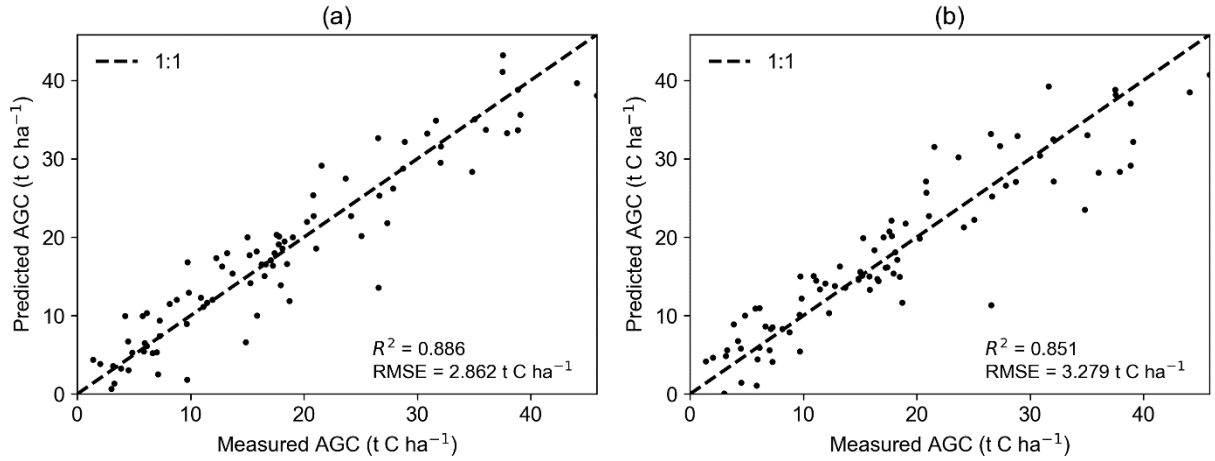

Fig. 6 Measured versus (a) multivariate, and (b) univariate, model predicted AGC

Table 5 Selected features

Rank	Feature	Coefficient
1	$Log(\overline{R/pan})$	-289.358
2	$Entropy(pan/B)$	3.356
3	$(\overline{B/C})^2$	-12.229
4	$Entropy(NIR1/Y)$	-2.114
5	$Std(R/Y)$	-0.278
6	$Entropy(B/NIR2)$	2.089
7	$Entropy(C/RE)$	-3.480
8	$(\overline{NIR1/NIR2})^2$	-7.984
9	$Entropy(G/NIR2)$	-0.500
10	$Std(C/G)$	131.919
11	$Std(Y/R)$	-93.340
12	$Entropy(R/G)$	14.267
13	$Std(NIR2/C)$	-3.487
14	$Entropy(R/pan)$	-9.351
15	$(\overline{Y/RE})^2$	-30.393

Accuracies of the multivariate and univariate AGC models are compared in Table 6. RMSE confidence intervals are derived from the LOOCV RMSE values. The relative RMSE (rRMSE) is the RMSE expressed as a percentage of the mean AGC ground truth.

Table 6 AGC regression evaluation

Description	Number of features	R^2	RMSE (t C ha^{-1})	5-95% RMSE CI (t C ha^{-1}) [a]	rRMSE (%) [a]
Multivariate	15	0.886	2.862	0.030 - 7.503	15.882
Univariate	1	0.851	3.279	0.247 - 9.194	18.196

[a] CI = confidence interval, rRMSE = relative RMSE

To better understand the influence of possibly noisy litter C estimates on AGC, the correlation between litter C and ABC was examined, and is plotted in Fig. 7. The AGC map produced by applying the multivariate model to the WV3 2017 image is shown in Fig. 8.

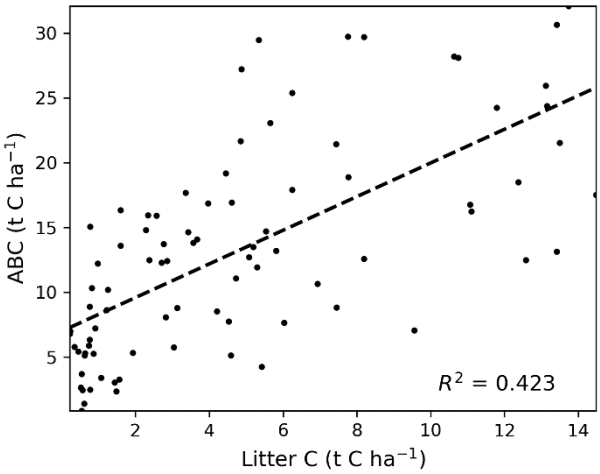

Fig. 7 Litter C and ABC correlation

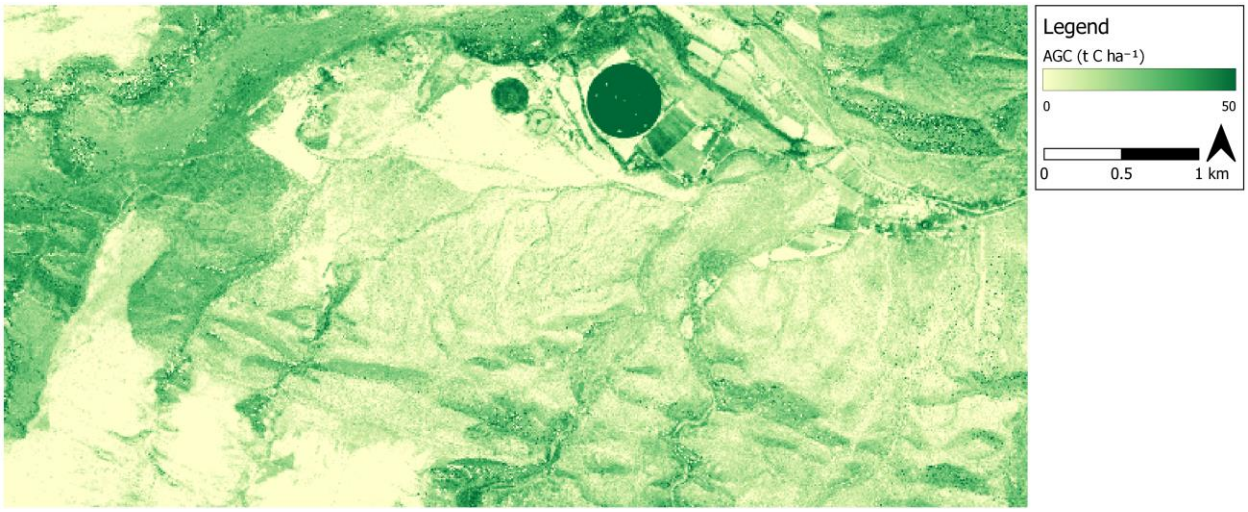

Fig. 8 Multivariate AGC map of the study area

5 Discussion

The performance of the multivariate model (Table 6) would be suitable for operational AGC estimation, and compares well with related studies.[18,22,33,52,53] Fig. 5 shows the effects of increasing complexity on model accuracy. As the number of selected features increases from one to 25, the explanatory power of the model increases with an associated increase in accuracy. At approximately 15 features, overfitting begins to occur as there insufficient data to represent the

underlying dimensionality. This is known as the "peaking phenomenon" and is the main motivation for feature selection.[26]

The 15 selected features (Table 5) include a combination of band ratio and texture features (and non-linear transformations thereof). Band ratios describe spectral properties of vegetation, such as greenness or relative shape of the red-edge. In remote sensing of biomass, texture features are known to complement spectral information.[19,21] A number of entropy and standard deviation features are also included in the selected set, confirming the usefulness of texture for predicting AGC. The majority of the AGC variance is however explained by the first selected feature $Log\left(\overline{R/pan}\right)$, a band ratio, as evidenced by the performance of the univariate model (R^2 = 0.851). The multivariate model provides a relatively small accuracy improvement over the univariate model (Table 6), at the cost of increased complexity. Compared to the univariate model, this complexity makes the multivariate model slower to compute and less easily transferred to new areas or images.

An investigation into possible error sources for outlying plots in Fig. 6 revealed that the worst performing field plot contained the second highest tree. It seems likely that the known lack of sensitivity to the tree height in optical data[21] contributed to this error, resulting in an underestimation of AGC for the large tree. There was not, however, a general pattern of AGC error being related to large trees in the rest of the data, and no other systemic sources of error were identified.

As the sampling plots are representative of thicket habitat only, estimated AGC values for other vegetation types may not be valid. There is a small amount of noise in the multivariate AGC output in some areas. This is due to the inclusion of texture features in the multivariate model which are sensitive to sudden changes in illumination caused by shadows, and sun-glint off bare

rock, for example. In practice, noise will be reduced when aggregating AGC values over extended areas.

It is known that biomass estimation accuracy reduces with sampling plot size and positional accuracy.[44,54] To counteract the effect of the relatively small sampling plot sizes (see e.g. Frazer et al. [44] for commonly used plot sizes), DGPS plot locations were post-processed to ±30 cm accuracy, and a 0.5 m DEM was produced from aerial imagery and used to orthorectify the Worldview-3 image. This allowed precise (sub-meter) location of sampling plots in the image and likely contributed to accuracy of the model.

The high correlation between field measured and model predicted AGC supports the use of the adopted sampling protocol ("standard operating procedure" (SOP))[46] for generating meaningful thicket AGC ground truth for remote sensing. The nesting approach[45] used in the SOP substantially reduces the number of plants requiring measurement. Further optimizations to the sampling protocol should be investigated by analyzing the AGC ground truth data. Small adjustments to the nesting approach can have a large impact on the cost of field work, which is currently a limiting factor in scaling up restoration efforts to the biome.[13,20]

The litter sampling approach is another aspect of the SOP that could benefit from further investigation, and possible modification. Litter carbon (C) forms a substantial component of AGC in intact subtropical thicket (e.g. the ratio of average litter C to average ABC was 0.433 for the intact stratum), and it is known that litter C forms a greater component of AGC in subtropical thicket than in other habitats.[8] Given the importance of litter C in thicket, one should ensure it is sampled accurately. It is questionable, however, if litter C sampled from four 0.5 m x 0.5 m plots is representative of litter C in the much larger containing plot. The combined area of the litter sample is 1 m^2, while the containing plot area is 100 m^2 or 400 m^2, depending on the stratum. It

seems likely that in heterogeneous thicket habitats the small area litter sample will be noisy relative to what is contained in the larger plot. This hypothesis would help explain the weak correlation (R^2 = 0.423) between litter C and ABC shown in Fig. 7.

Two-step or multi-scale approaches are seeing increasing use for regional and national scale vegetation mapping.[24,55,56] In these approaches, small scale field ground truth is extrapolated to larger spatial extents using regression modelling with VHSR imagery. In a second step, medium resolution imagery is used with the extrapolated ground truth to derive vegetation maps over much larger spatial extents. The multi-scale approach achieves a good compromise between the costs and benefits of the field, VHSR and medium resolution worlds, and represents a promising avenue for biome-wide AGC mapping in thicket. The simplicity and accuracy of the univariate model lends itself to use in multi-scale studies investigating extension of AGC mapping in thicket. In contrast to the multivariate model, the absence of any VHSR dependent texture features in the univariate model means it can easily be applied to freely available medium resolution optical imagery, such as Landsat[57] or Sentinel-2[58], with coverage of the biome.

As it stands, the remote sensing approach is valid for inclusion under the "AR-ACM0003" CDM (Clean Development Mechanism) carbon accounting methodology[59], which is accepted under the preferred verification standards for the South African national carbon tax.[16,17] The AGC map of the study area (Fig. 8) serves as a baseline carbon assessment for future restoration or other studies in the area.

6 Conclusion

A snapshot regression model using 15 band ratio, vegetation index and textural features from WorldView-3 imagery provided good AGC prediction accuracy (R^2 = 0.886 and RMSE = 2.862 t C ha^{-1}). While remote sensing of plant structure with e.g. LiDAR or SAR data could prove

to be effective for mapping thicket AGC, challenges with those approaches precluded their evaluation in this study. The results of this, and the Nyamugama & Kakembo[27] study, demonstrate the efficacy of multi-spectral imagery for accurate remote sensing of AGC in thicket. The models developed here provide an important foundation for future work investigating the spatial extension of AGC maps to the thicket biome. The univariate model is well-suited to multi-scale approaches, which allow the extension of vegetation maps to regional and national extents,[24,56] and reduce the need for costly VHSR imagery and field work.[55] AGC maps will assist in the planning and monitoring of restoration, verification of stored carbon and understanding of thicket ecosystem dynamics. The presented remote sensing approach provides a starting point for reducing the monitoring costs for rehabilitation carbon projects in the thicket biome – an essential component to enable broader-scale restoration.

Table of Figures

Disclosure

There are no conflicts of interest to declare.

Acknowledgments

This research was funded by the GEF-5 SLM project ("Securing multiple ecosystems benefit through sustainable land management (SLM) in the productive but degraded landscapes of South Africa") via the UNDP and Rhodes University. WorldView-3 imagery was acquired courtesy of the DigitalGlobe Foundation. The authors thank Rebecca and Mike Powell for ongoing discussions that helped shape and direct this study, Marius van der Vyver for providing allometric models and guidance on their use, Melloson Allen from Living Lands, and Hanton Windvogel, who assisted with field work, the Centre for Geographical Analysis (CGA) at Stellenbosch University for help with the acquisition and processing of NGI aerial imagery, and the departments of geography at Stellenbosch and Rhodes universities for the loan of DGPS devices. We also thank the anonymous reviewers for their constructive suggestions that have much improved the manuscript.

References

1. J. H. J. Vlok and D. I. W. Euston-Brown, “The patterns within, and the ecological processes that sustain, the subtropical thicket vegetation in the planning domain of the Subtropical Thicket Ecosystem Planning (STEP) project,” Terrestrial Ecology Research Unit, University of Port Elizabeth, Port Elizabeth (2002).
2. R. M. Cowling, Ô. ProcheÕ, and J. H. J. Vlok, “On the origin of southern African subtropical thicket vegetation,” South African J. Bot. **71**(1), 1–23 (2005) [doi:10.1016/S0254-6299(15)30144-7].
3. A. J. Mills et al., “Effects of goat pastoralism on ecosystem carbon storage in semiarid thicket, Eastern Cape, South Africa,” Austral Ecol. **30**, 797–804 (2005) [doi:10.1111/j.1442-9993.2005.01523.x].
4. R. G. Lechmere-Oertel, G. I. H. Kerley, and R. M. Cowling, “Patterns and implications of transformation in semi-arid succulent thicket, South Africa,” J. Arid Environ. **62**(3), 459–474 (2005) [doi:10.1016/j.jaridenv.2004.11.016].
5. J. W. Lloyd, E. C. Van den Berg, and A. R. Palmer, “Patterns of transformation and degradation in the Thicket Biome, South Africa,” Terrestrial Ecology Research Unit, University of Port Elizabeth, Port Elizabeth (2002).
6. M. Thompson et al., “Mapping grazing-induced degradation in a semi-arid environment: A rapid and cost-effective approach for assessment and monitoring,” Environ. Manage. **43**, 585–596 (2009) [doi:10.1007/s00267-008-9228-x].
7. A. Sigwela et al., “The impact of browsing-induced degradation on the reproduction of subtropical thicket canopy shrubs and trees,” South African J. Bot. **75**(2), 262–267, Elsevier B.V. (2009) [doi:10.1016/j.sajb.2008.12.001].

8. R. G. Lechmere-Oertel et al., “Litter dynamics across browsing-induced fenceline contrasts in succulent thicket, South Africa,” South African J. Bot. **74**(4), 651–659 (2008) [doi:10.1016/j.sajb.2008.04.002].

9. M. L. Van der Vyver et al., “Spontaneous return of biodiversity in restored subtropical thicket: *Portulacaria afra* as an ecosystem engineer,” Restor. Ecol. **21**(6), 736–744 (2013) [doi:10.1111/rec.12000].

10. A. J. Mills and R. M. Cowling, “Rate of carbon sequestration at two thicket restoration sites in the Eastern Cape, South Africa,” Restor. Ecol. **14**(1), 38–49 (2006) [doi:10.1111/j.1526-100X.2006.00103.x].

11. C. Clarke, S. Shackleton, and M. Powell, “Climate change perceptions, drought responses and views on carbon farming amongst commercial livestock and game farmers in the semiarid Great Fish River Valley, Eastern Cape province, South Africa,” African J. Range Forage Sci. **29**(1), 13–23 (2012) [doi:10.2989/10220119.2012.687041].

12. A. J. Mills et al., “Investing in sustainability. Restoring degraded thicket, creating jobs, capturing carbon and earning green credit,” Climate Action Partnership, Cape Town and Wilderness Foundation, Port Elizabeth (2010).

13. M. J. Powell, “Restoration of degraded subtropical thickets in the Baviaanskloof Megareserve, South Africa,” Master’s Thesis, Rhodes University, Department of Environmental Science (2009).

14. A. J. Mills et al., “Assessing costs, benefits, and feasibility of restoring natural capital in Subtropical Thicket in South Africa,” in Restoring Natural Capital: Science, Business and Practice (The Science and Practice of Ecological Restoration Series) **02**, J. Aronson, S. Milton, and J. Blignaut, Eds., pp. 179–187, Island Press, Washington DC (2007).

15. C. Marais, R. M. Cowling, and M. Powell, “Establishing the platform for a carbon sequestration market in South Africa: The Working for Woodlands Subtropical Thicket Restoration Programme,” in XIII World Forestry Congress (October), pp. 1–13, Buenos Aires (2009).

16. J. Reeler, “Carbon market methodological options for restoration of subtropical thicket: A review of options and implications for restoration action for the GEF5 SLM project Baviaanskloof component,” Rhodes University, GEF-5 SLM, Makhanda (2018).

17. T. Knowles et al., “Development of potential verification standards and methodologies for carbon offset projects in the AFOLU sector in South Africa,” Department of Environmental Affairs, Pretoria (2015).

18. C. Eisfelder, C. Kuenzer, and S. Dech, “Derivation of biomass information for semi-arid areas using remote-sensing data,” Int. J. Remote Sens. **33**(9), 2937–2984 (2012) [doi:10.1080/01431161.2011.620034].

19. D. Lu, “The potential and challenge of remote sensing-based biomass estimation,” Int. J. Remote Sens. **27**(7), 1297–1328 (2006) [doi:10.1080/01431160500486732].

20. M. L. Van der Vyver and R. M. Cowling, “Aboveground biomass and carbon pool estimates of Portulacaria afra (spekboom)-rich subtropical thicket with species-specific allometric models,” For. Ecol. Manage. **448**, 11–21 (2019) [doi:10.1016/j.foreco.2019.05.048].

21. D. Lu et al., “A survey of remote sensing-based aboveground biomass estimation methods in forest ecosystems,” Int. J. Digit. Earth **9**(1), 63–105 (2016) [doi:10.1080/17538947.2014.990526].

22. I. Ali et al., “Review of machine learning approaches for biomass and soil moisture

retrievals from remote sensing data," Remote Sens. **7**(12), 16398–16421 (2015) [doi:10.3390/rs71215841].

23. R. W. Kulawardhana, S. C. Popescu, and R. A. Feagin, "Airborne lidar remote sensing applications in non-forested short stature environments: A review," Ann. For. Res. **60**(1), 127–150 (2017) [doi:10.15287/afr.2016.719].
24. R. Mathieu et al., "A radar- and LiDAR-based earth observation system for monitoring savanna woody structure in southern Africa," Biodivers. Ecol. **6**, 355–359 (2018) [doi:10.7809/b-e.00345].
25. P. M. Mather and M. Koch, *Computer processing of remotely-sensed images*, John Wiley & Sons, Chichester (2011) [doi:10.1002/9780470666517].
26. A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: A review," IEEE Trans. Pattern Anal. Mach. Intell. **22**(1), 4–37 (2000) [doi:10.1109/34.824819].
27. A. Nyamugama and V. Kakembo, "Estimation and monitoring of aboveground carbon stocks using spatial technology," S. Afr. J. Sci. **111**(9/10), 1–7 (2015) [doi:10.17159/sajs.2015/20140170].
28. D. Harris, J. Vlok, and A. Van Niekerk, "Regional mapping of spekboom canopy cover using very high resolution aerial imagery," J. Appl. Remote Sens. **12**(04), 1 (2018) [doi:10.1117/1.JRS.12.046022].
29. National Geo-spatial Information, "National aerial photography and imagery programme," Department of Rural Development and Land Reform (DRDLR), 2012, <http://www.ngi.gov.za/index.php/what-we-do/aerial-photography-and-imagery> (accessed 22 May 2012).
30. D. Harris and A. Van Niekerk, "Radiometric homogenisation of aerial images by

calibrating with satellite data," Int. J. Remote Sens. **40**(7), 2623–2647, Taylor & Francis (2019) [doi:10.1080/01431161.2018.1528404].

31. D. Harris and A. Van Niekerk, "Feature clustering and ranking for selecting stable features from high dimensional remotely sensed data," Int. J. Remote Sens. **39**(23), 8934–8949 (2018) [doi:10.1080/01431161.2018.1500730].
32. D. B. Hoare et al., "Albany Thicket Biome," in The vegetation of South Africa, Lesotho and Swaziland, L. Mucina and M. C. Rutherford, Eds., pp. 540–567, SANBI, Pretoria (2006).
33. N. Clerici et al., "Estimating aboveground biomass and carbon stocks in periurban Andean secondary forests using very high resolution imagery," Forests **7**(7) (2016) [doi:10.3390/f7070138].
34. F. E. Fassnacht et al., "Review of studies on tree species classification from remotely sensed data," Remote Sens. Environ. **186**, 64–87 (2016) [doi:10.1016/j.rse.2016.08.013].
35. A. Skowno and D. I. W. Euston-Brown, "BMR Vegetation 2006," SANBI BGIS, 2006, <http://bgis.sanbi.org/SpatialDataset/Detail/249> (accessed 30 October 2019).
36. Agisoft, "Agisoft Metashape," 2019, <https://www.agisoft.com/> (accessed 30 October 2019).
37. R. Richter, "Correction of atmospheric and topographic effects for high spatial resolution satellite imagery," Int. J. Remote Sens. **18**(5), 1099–1111 (1997) [doi:10.1080/014311697218593].
38. PCI Geomatics, "PCI Geomatica," 2021, <https://www.pcigeomatics.com/software/geomatica/professional> (accessed 10 June 2021).

39. A. Ghosh and P. K. Joshi, “Assessment of pan-sharpened very high-resolution WorldView-2 images,” Int. J. Remote Sens. **34**(23), 8336–8359 (2013) [doi:10.1080/01431161.2013.838706].

40. P. Cheng and C. Chaapel, “Pan-sharpening and geometric correction: WorldView-2 satellite,” in GEOInformatics, pp. 30–33 (2010).

41. C. Bolus et al., “Above ground carbon baseline assessment, using species specific allometric models, across three degradation classes in Baviaanskloof Spekboom Thicket,” Rhodes University, GEF-5 SLM, Makhanda (2020).

42. J. H. J. Vlok, “Guidelines for spekboom and veld restoration on sewefontein and tchuganoo,” Rhodes University, GEF-5 SLM, Makhanda (2017).

43. F. E. Fassnacht et al., “Importance of sample size, data type and prediction method for remote sensing-based estimations of aboveground forest biomass,” Remote Sens. Environ. **154**(1), 102–114 (2014) [doi:10.1016/j.rse.2014.07.028].

44. G. W. Frazer et al., “Simulated impact of sample plot size and co-registration error on the accuracy and uncertainty of LiDAR-derived estimates of forest stand biomass,” Remote Sens. Environ. **115**(2), 636–649 (2011) [doi:10.1016/j.rse.2010.10.008].

45. S. Lackmann, “Good practice in designing a forest inventory,” Capacity Development for REDD+, Quito (2011).

46. C. Bolus et al., “Standard operating procedure (SOP) for above ground carbon baseline field sampling,” Rhodes University, GEF-5 SLM, Makhanda (2018).

47. P. Blauensteiner et al., “On colour spaces for change detection and shadow suppression,” in Computer Vision Winter Workshop 2006, O. Chum and V. Franc, Eds., pp. 1–6, Czech Pattern Recognition Society, Telc, Czech Republic (2006).

48. J. B. Campbell and R. H. Wynne, *Introduction to remote sensing*, The Guilford Press, New York (2011).

49. A. R. Huete, "A soil-adjusted vegetation index (SAVI)," Remote Sens. Environ. **25**(3), 295–309 (1988) [doi:10.1016/0034-4257(88)90106-X].

50. R. Trias-Sanz, G. Stamon, and J. Louchet, "Using colour, texture, and hierarchial segmentation for high-resolution remote sensing," ISPRS J. Photogramm. Remote Sens. **63**(2), 156–168 (2008) [doi:10.1016/j.isprsjprs.2007.08.005].

51. C. M. Bishop, *Neural networks for pattern recognition*, Oxford University Press, New York (2003).

52. S. Eckert, "Improved forest biomass and carbon estimations using texture measures from WorldView-2 satellite data," Remote Sens. **4**(4), 810–829 (2012) [doi:10.3390/rs4040810].

53. M. Singh et al., "Mapping above-ground biomass in a tropical forest in Cambodia using canopy textures derived from Google Earth," Remote Sens. **7**(5), 5057–5076 (2015) [doi:10.3390/rs70505057].

54. M. Réjou-Méchain et al., "Upscaling forest biomass from field to satellite measurements: Sources of errors and ways to reduce them," Surv. Geophys. **40**(4), 881–911 (2019) [doi:10.1007/s10712-019-09532-0].

55. M. González-Roglich and J. J. Swenson, "Tree cover and carbon mapping of Argentine savannas: Scaling from field to region," Remote Sens. Environ. **172**, 139–147 (2016) [doi:10.1016/j.rse.2015.11.021].

56. M. Immitzer et al., "Fractional cover mapping of spruce and pine at 1 ha resolution combining very high and medium spatial resolution satellite imagery," Remote Sens.

Environ. **204**, 690–703, Elsevier (2018) [doi:10.1016/j.rse.2017.09.031].

57. G. Schmidt et al., “Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) Algorithm Description,” USGS, Reston (2013) [doi:10.3133/ofr20131057].

58. ESA, “Sentinel-2 User Handbook,” 2015, <https://sentinel.esa.int/documents/247904/685211/Sentinel-2_User_Handbook> (accessed 27 July 2018).

59. CDM, “A/R Large-scale consolidated methodology: Afforestation and reforestation of lands except wetlands,” UNFCCC (2013).

Biographies

Dugal Harris (PhD Stellenbosch University) is a freelance software engineer specializing in earth observation, machine learning and geospatial applications. His research interests lie in the use of remote sensing to inform responses to environmental and socio-economic issues. He has a diverse work and research background, with experience in the fields of software engineering, image processing and machine learning.

Cosman Bolus (MSc Environmental Science Rhodes University, BSc Hons Botany UCT) is a freelance environmental scientist and social-ecologist with a focus on ecological carbon and regenerative agriculture projects. He has an interest and extensive work experience in carbon and biodiversity baseline assessments, vegetation restoration methodologies (in Thicket, Fynbos and the Grassland biomes) and natural resource management (NRM) in water catchments on communal lands. He is the founder of Sustainable Landscape Solutions and is also an associate of the Rhodes Restoration Research Group (RRRG).

James Reeler (MSc University of Cape Town) is the bioenergy project manager for WWF South Africa. He has more than ten years' experience working with climate change issues covering both mitigation and adaptation, and ranging from local impacts to international policy.