Regional mapping of spekboom canopy cover using very high resolution aerial imagery

Dugal Harris,a\* Jan Vlok,b Adriaan van Niekerk,a

aStellenbosch University, Department of Geography and Environmental Studies, Stellenbosch, South Africa, 7602

bNelson Mandela Metropolitan University, Department of Botany, Port Elizabeth, South Africa, 6031

**Abstract**. Very high resolution canopy cover maps of spekboom are required to assist with the restoration of degraded habitat in the Little Karoo, a large semi-arid region in South Africa. Variations in habitat and level of degradation, in addition to radiometric variations in the imagery, make spekboom mapping at a regional scale a challenging problem. In this article, we present a per-pixel classification approach for canopy cover mapping of spekboom using multi-spectral 0.5 m resolution aerial imagery. The imagery was radiometrically homogenized with a technique that uses satellite data to convert digital numbers to estimated surface reflectance values. A feature selection procedure that is robust to redundancy was applied in order to select an informative feature subset from a typical set of spectral, textural and vegetation index features. Support vector machine (SVM), random forest, decision tree, k-nearest neighbor (kNN) and Bayes normal classifiers were evaluated against labeled pixel data and canopy cover ground truth acquired at 20 field sites. The results showed that all the classifiers, except the Bayes normal classifier, performed well. The decision tree produced the best results (mean absolute canopy cover error of 5.85% with a standard deviation of 4.65%).

**Keywords**: very high resolution, aerial imagery, vegetation mapping, canopy cover, radiometric calibration, feature redundancy.

**\***Dugal Harris**,** E-mail: [dugalh@gmail.com](mailto:dugalh@gmail.com)

# Introduction

Spekboom (*Portulacaria afra*) is an evergreen succulent tree with a dense canopy of small fleshy leaves that can grow up to 2.5 m in height. It occurs in the subtropical thicket biome in the semi-arid Eastern Cape and Little Karoo regions of South Africa.1 “The Subtropical Thicket habitat types are most easily recognized by the occurrence of woody trees, spinescent shrubs and a relative abundance of succulents. When occurring as solid stands the vegetation can form impenetrable dense thickets, but this is uncommon in the Little Karoo. In most of the area the Subtropical Thicket vegetation occurs as discrete bush-clumps, usually in a matrix of Succulent Karoo vegetation”.1 While spekboom tolerates browsing by indigenous herbivores, it is highly susceptible to over-browsing by goats.2–5 Poorly managed goat grazing has transformed thicket over much of its range into sparsely scattered thicket clumps, isolated trees and a covering of herbs.5

The benefits of restoring degraded thicket habitat are evident from a number of perspectives. Spekboom is unusually effective at storing carbon for an arid region plant.5 Subtropical thicket furthermore provides an important source of food for many herbivores, including domesticated livestock.1,6 The re-establishment of spekboom in degraded areas will reduce erosion and flood severity and improve water quality.7,8 The restoration of spekboom is also attractive from an employment perspective, since the restoration process could potentially create thousands of jobs in impoverished areas if implemented on a large scale. Currently, the most practical option for thicket ecosystem restoration is through the planting of spekboom cuttings.4,9 Spekboom is a keystone species and facilitates the creation of a favorable environment for the spontaneous recruitment of other plants.10,11

Spekboom canopy cover maps are required for assisting in the restoration process. There is a need for greater accuracy and repeatability than that provided by field-based mapping techniques. Field mapping is time consuming and costly and is not practical over large areas.12–14 Manual field mapping is confounded by the density (inaccessibility), heterogeneous nature and complex growth forms of the subtropical thicket biome.13

Thompson et al.Thompson et al.6 conducted a general degradation mapping study of the biomes occurring in the Little Karoo. A 1:50000 vegetation map, developed by Vlok, Cowling and Wolf1, was used to delineate different habitats so they could be treated separately. A coarse three-level degradation classification of subtropical thicket was derived by thresholding 250 m resolution MODIS normalized difference vegetation index (NDVI) data. The study was successful at estimating three degradation levels (intact, moderate and severe) of spekboom thicket at the 250 m MODIS resolution.

There is an initiative to involve private land owners in subtropical thicket restoration in order to broaden its impact.15,16 Spekboom is to be planted in stand sizes as small as three hectares on these lands. VHR imagery is required to produce maps of sufficient spatial detail for accurately monitoring canopy cover in these small planting stands. Spekboom often occurs in small clumps amongst a complex and varying mosaic of other vegetation. Moderate to low resolution imagery suffers from spectral mixing between Spekboom, other vegetation and soil in this heterogeneous biome6. VHR imagery is therefore also necessary to facilitate fine scale discrimination of spekboom from surrounding soil and vegetation.. To achieve sufficient accuracy for restoration monitoring and carbon storage estimations, it is necessary to estimate canopy cover in finer detail than the three levels of degradation and 250 m resolution used in Thompson et al.Thompson et al.6.

Multi-spectral VHR imagery has been successfully used for vegetation mapping in a number of studies. A combination of spectral band, vegetation index, band ratio and textural features are commonly used to provide informative measures capable of distinguishing vegetation classes.17–21 Object-based approaches to image classification, where homogenous image objects are first generated through segmentation and then classified as a whole, have become popular in vegetation studies.18,20,22 These approaches are often favored for VHR imagery17–22 because they are potentially able to better exploit the additional spatial information and deal with unwanted variation when compared to the more traditional per-pixel approach.18,22 The segmentation problem is, however, recognized as being poorly posed, requiring manual adjustment of parameters and being difficult to solve.23 Per-pixel classification provided good and useful mapping accuracy in a number of studies24–26 and is a simpler and faster method, not requiring user specification of algorithms and associated parameters.

A variety of supervised approaches are used for classifying features derived from VHR imagery. Some authors found the Bayes normal (maximum likelihood (ML)) classifier to adequately model their class distributions.24–26 Others adopted more sophisticated approaches such as SVMs18 and neural networks.17,21 Algorithms implemented in the eCognition software package,27 such as the fuzzy and hierarchical approaches, are also frequently used for VHR image classification.19,20,22

As the number of features increases, the amount of data required to adequately represent class distributions in the increased feature space increases exponentially. This is known as the “curse of dimensionality”.28 For finite training samples, increasing the features beyond a certain point results in overtraining and a decrease in the classifier’s ability to generalize. This “peaking phenomenon”29 makes it necessary to apply feature selection to reduce the size of the feature-set to a salient minimum in order to achieve an accurate classification. Feature selection by ranking, based on some separability or importance measure of individual features, is frequently used.17,19,20 While fast, feature ranking is known to be sub-optimal for feature spaces containing redundancy.30 Ghosh and Joshi18 used recursive feature elimination (also known as backward elimination) – a greedy search technique to select informative features. Of the reviewed studies, Ghosh and Joshi18 were the only ones to use a feature selection method that considers the effect of feature redundancy by evaluating features in combination.

The majority of the reviewed vegetation mapping studies were applied to small areas, typically covered by a single satellite image.18–20,22,24,26 Radiometric corrections are sometimes not applied in small study areas24 or partially handled using conversion to top of atmosphere radiance.19,26 These corrections do not compensate for varying atmospheric and bi-directional distribution function (BRDF) effects, characteristic of datasets containing hundreds or thousands of aerial images.

In this paper, we present a method for mapping spekboom canopy cover at a spatial resolution of 0.5 m. A total of 2228 multi-spectral aerial images, acquired over multiple days from 22 January to 8 February 2010, were used as input. Radiometric variations due to atmospheric and BRDF effects in the images were reduced using a simple yet effective technique for homogenizing the digital numbers to approximate surface reflectance. This not only allowed for the application of a single classification algorithm to the entire set of images, but also provides the possibility of extending the presented mapping technique spatially and temporally. An informative feature subset was selected from a typical set of spectral band, band ratio, vegetation index and textural features using a novel feature selection method that is robust to redundancy typically found in high-dimensional feature-sets. The selected features were used to evaluate a set of candidate classifiers.

# Data

1

|  |
| --- |
| 2 |

1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

3

31

2

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |

4

**3**

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

# Methods and Experiments

## Radiometric Homogenization

While the imagery provided by NGI is orthorectified, no radiometric corrections were applied to it. The NGI imagery contains variations due to BRDF and atmospheric effects, which makes it poorly suited to quantitative remote sensing techniques. The imagery was consequently radiometrically homogenized through the application of a surface reflectance estimation technique. This technique corrects for coarse scale atmospheric and BRDF effects using a well-calibrated, concurrent and collocated surface reflectance satellite image as a reference. We used a MODIS MCD43A4 composite image for the period of 25 January 2010 to 9 February 2010 for this purpose. This image has a 500 m resolution and contains nadir BRDF-adjusted reflectance data composited from the best values over a 16-day period. While Sentinel-232 or Landsat33 surface reflectance could also serve as reference data, no cloud-free imagery concurrent (or near-concurrent) to the aerial imagery was available from these sources. The relative spectral responses (RSR’s) of the DMC and corresponding MODIS bands are shown in Fig. 2. Radiometric correction is important as it allows accurate snapshot mapping of large spatial extents and provides the possibility of repeating the canopy-cover mapping to evaluate restoration progress.



Fig. 2 MODIS and DMC RSR’s

## Mapping Methodology

The image resolution of 0.5 m, combined with the tendency of spekboom to grow in continuous stands, meant that there was little spectral mixing and that pixels covering spekboom were relatively pure. This supported a per-pixel classification approach to distinguish spekboom from the surrounding vegetation. The pixel-based approach also ensured that the complexities associated with segmentation could be avoided. The fractional canopy cover was determined as the portion of pixels classified as spekboom over an area of interest.

Given the large number of images, computation time was an important consideration in the formulation of our method. Radiometric homogenization and classification software tools were developed using the GDAL34 and OpenCV35 software libraries. Careful consideration was given to computational efficiency in the selection of features and classification algorithm.

|  |
| --- |
|  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

## Features

A set of 46 features that would aid in describing the visual characteristics of spekboom were evaluated. The set included a typical combination of spectral features, vegetation indices and texture features. Similar features have been used in Li et al.36 and Trias-Sanz, Stamon and Louchet.37 The features can be grouped into two broad categories: per-pixel and sliding window features. The per-pixel features are found with the spectral information from only one pixel, while the sliding window features are found from a statistic of the pixels inside a small local neighborhood. While the spectral resolution of the VHR imagery is poor, the spatial resolution enables a description of the vegetation structure and spatial patterns, which is not possible with lower resolution satellite imagery. Texture features are a popular way of encapsulating spatial and structural information. Measures of vegetation texture are sensitive to shadow variations, an unavoidable phenomenon in aerial imagery caused by the long flight times and varying sun angle. Nevertheless, texture is recognized as an important feature in biomass estimation in complex habitats.14,38 The sliding window features were included to exploit this source of information when distinguishing the classes.

Although the imagery was calibrated to surface reflectance, it was done at a coarse spatial scale and fine resolution radiometric variations were not taken into account. A normalized color space was consequently included in the features to reduce intensity variations not removed by the surface reflectance corrections. Color is captured by the relative amounts of the raw color bands rather than their absolute values. Normalized color features are defined as:39

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  | (1) |

where are the raw R, G, B and NIR band values and is the band number. The denominator normalizes for intensity.

Green, living vegetation absorbs light in the photosynthetically active radiation region of the spectrum, which corresponds to the red band. There is a sharp transition from absorption to reflection around 700 nm40 which is captured by the DMC R and NIR wavelength ranges (see Fig. 2). Vegetation is highly reflective in the near-infrared band as the energy in these wavelengths is insufficient for photosynthesis and potentially harmful due to its heating effects. Various vegetation indices exploit these spectral properties. The ratio vegetation index (RVI) is given by:

|  |  |  |
| --- | --- | --- |
|  |  | (2) |

It has a range of zero to infinity and increases as the vegetation becomes denser and photosynthetically more active.41 The well-known NDVI is defined as

|  |  |  |
| --- | --- | --- |
|  |  | (3) |

NDVI is limited to the range -1 to 1 and contains the same information as RVI, but is easier to visualize and interpret due to its limited range. Both indices are unaffected by intensity changes.

The tasseled cap transform is a linear transform of the raw band feature space to a new orthogonal co-ordinate system, similar to a principal component transform. It was designed for agricultural wheat classification and was intended to reduce variability in soil and wheat classes by removing variation due to topography, sun angle and wheat growth stage.42 The tasseled cap transform was approximated in this study by using a principal component transform derived from the variance of the spekboom class. The first component was aligned with spekboom variation rather than wheat variation, as in the original tasseled cap transform. As it is simply a rotation of the raw band space, it is more useful as a dimensionality reduction technique (similar to principal components analysis (PCA)) than an extractor of novel features. The principal components of the normalized colors of Equation (1) were also included as features in the classification process.

Entropy is a statistic that describes the amount of randomness in a variable. It was included in our feature-set as a texture feature to describe complexity in the local neighborhood of a sliding window. The entropy of the values in the image window is defined as:37

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  | (4) |

where is the probability in the histogram bin of . A total of 256 bins were used in all cases.

In addition to the entropy, the median and the four central moment features (mean, standard deviation, skewness and kurtosis) of Li et al.36 were included as sliding window features. The first principal component, RVI, NDVI and normalized green channel were all used as inputs to the sliding window feature-set. The complete feature-set and their labels are listed in Table 4. A sliding window size of five was selected using a cross-validated grid search, with the accuracy of a naïve Bayes classifier trained on the EntropyPc1 feature as the performance criterion. This size seemed sensible as it is comparable to that of a small spekboom clump.

**Table 4** Features

|  |  |  |
| --- | --- | --- |
| No. | Name | Description |
| 1 | R | Red |
| 2 | G | Green |
| 3 | B | Blue |
| 4 | NIR | Near-infrared |
| 5 | rN | Normalized R |
| 6 | gN | Normalized G |
| 7 | bN | Normalized B |
| 8 | nirN | Normalized NIR |
| 9 | NDVI | Normalized difference vegetation index |
| 10 | RVI | Ratio vegetation index |
| 11–14 | tc1–4 | Tasseled cap components |
| 15–18 | pc1–4 | Principal components of raw bands |
| 19–22 | nc1–4 | Principal components of normalized bands |
| 23–26 | Entropy## | Sliding window entropy of pc1, RVI, NDVI and gN |
| 27–30 | Std## | Sliding window standard deviation of pc1, RVI, NDVI and gN |
| 31–34 | Mean## | Sliding window mean of pc1, RVI, NDVI and gN |
| 35–38 | Median## | Sliding window median of pc1, RVI, NDVI and gN |
| 39–42 | Skewness## | Sliding window skewness of pc1, RVI, NDVI and gN |
| 43–46 | Kurtosis## | Sliding window kurtosis of pc1, RVI, NDVI and gN |

## Feature Selection

The bands of the imagery have significant spectral overlap43 and consequently are highly correlated. Given that the bands are the source data for all the derived features, the derived feature definitions also contain inter-dependencies. A number of authors have noted that feature redundancy can cause instability and sub-optimality in selected features when traditional approaches (such as ranking, forward selection and backward elimination) are used.30,44–46 Redundancy can be reduced by using a feature extraction approach such as PCA, but requires computation of the full feature-set and is not practical in computationally demanding applications such as ours.

A feature selection method, called feature clustering and ranking, was used to select relevant features in the presence of redundancy. The approach is based on Harris47 but uses hierarchical clustering instead of affinity propagation to group redundant features. It is described as follows:

1. Perform average-linkage hierarchical clustering48 of the feature set using the correlation coefficient as the dissimilarity metric.
2. Select a dissimilarity threshold at which to extract a natural number of clusters containing high correlation by visual inspection of the dendrogram.
3. Rank each cluster’s importance by finding the value of a relevance criterion for each individual feature and then finding the median of the feature relevance values in the cluster.
4. Select a single feature from each of the *N* clusters with the best importance scores.

The number of clusters, *N*, was chosen using a grid search with the final classifier accuracy as performance measure. In this study, the accuracy of a naïve Bayes classifier was used as the feature relevance criterion. The naïve Bayes criterion makes no assumption about the form of the class distributions and can thus provide a generic measure of separability. It is simple, fast and recognized as being accurate for a variety of problems.49 To avoid biased accuracy estimates, all classifier accuracy evaluations for feature relevance or selection of *N*, were done on unseen test data using a ten fold cross validation.28 The cluster-ranking method has the advantages of being quick and allowing hand-picking of the single features that represent each cluster. The flexibility to choose features enables the user to favor those features that are fastest to compute, or perhaps to choose those features that are more readily understood. The method was applied to the labeled pixel data.

## Classification and Canopy-Cover Estimation

The decision tree, random forest, SVM, Bayes normal and k-nearest-neighbor (kNN) classifiers were evaluated in this study. A decision tree is a tree of binary decision nodes based on thresholds of different features. Data is recursively split at each branch node until a terminal representing a class label is reached.50 Training is performed by a greedy procedure, which iteratively adds nodes and selects features producing the best split for each node. Criteria used for choosing the best feature at each node include the information content, node purity and Fisher’s criterion.29 Overtraining is a concern and trees can be pruned in a post-training step to reduce variance. Decision trees are known for their speed of execution and ease of interpretation. Node decisions can help provide insight into the problem. Decisions are usually binary and based on a single feature. As a result, the decision boundary is comprised of stepwise sections parallel to the feature axes and is at best an approximation of the optimal boundary.29 Decision trees are flexible and broadly applied as they are non-parametric (i.e. they make no assumption about the form of class distributions) and can deal with categorical as well as continuous variables.50

Random forests are classifiers that use bootstrapped aggregation (bagging)51 of a large collection of decision tree classifiers. Each tree is trained on a bootstrapped version of the dataset and the decision feature for each node is selected from a random subset of the full feature set.52 The bootstrapping and random feature subsets help introduce variation amongst the base tree classifiers. The uncorrelated decision trees, in combination, have greater predictive power than any single one. Importantly, a random forest is not prone to overtraining. Random forests are also robust to mislabelled training data. Both training and execution demand a moderate amount of computation time. The two main parameters for tuning a random forest are the number of trees and the number of features considered for each node.

Ground-breaking and widespread pattern recognition work has been done with the SVM.36,44,46,53 The SVM was initially defined as the two-class linear decision boundary that maximized the distance to the nearest objects, called “support vectors”.54 The decision boundary is determined only by the support vectors, not directly by features or generative descriptions of class distributions. The SVM minimizes the Vapnik-Chervonenkis dimension, a measure of the complexity of the classifier. This is an important property of the SVM and explains how it effectively adapts its complexity to the data, is robust to overtraining and performs well in high-dimensional feature spaces. The original formulation was extended to the case of overlapping multi-class problems using a penalty term with user-defined multiplier *C*, which punishes class overlap. Using the kernel trick, the linear SVM was further extended to allow modeling of non-linear decision boundaries.54 Different kernels such as polynomials or radial basis functions (RBF) may be chosen to suit the given problem. In kernel form, the SVM can be considered a non-parametric classifier. In our evaluation, an RBF kernel was used for the SVM classifier. The training is done by using a computationally demanding quadratic optimization problem. However, execution is fast as it only requires an evaluation of the kernel function for the support vector – object vector pairs.29

The Bayes normal classifier, sometimes referred to as the Maximum Likelihood (ML) classifier, assumes that the classes are normally distributed. Mean and covariance parameters are estimated for each class from the data, usually with the ML criterion. Bayes’ rule is then used to define the decision boundary.55

The kNN classifier labels test objects by finding the mode of classes of the closest k training objects.28 Any distance metric can be used for finding neighbors, but the Euclidean distance measure is prevalent and was used in our study. This classifier is a useful benchmark as it almost always performs reasonably well, requires only one parameter and is non-parametric.29 It requires finding distances to the full training set, which can slow execution for large datasets.

User supplied tuning parameters for the classifiers were found with cross-validated grid searches. Table 5 details the parameter values selected for each classifier. Descriptions of the parameters can be found in the OpenCV documentation.56

Morphological operators57 were applied as a post-processing step to the classifier produced maps to remove noise and smooth boundaries. Assuming that the majority of spekboom plants were big enough to cover more than one pixel, a morphological opening was applied to remove isolated spekboom pixels. Following this, spurious wrinkles and holes in the spekboom boundaries were removed with a morphological closing operation; the assumption being that spekboom typically grows in solid clumps and any real gaps in these clumps would be more than a pixel wide. These operations can be seen as a way of further incorporating spatial context into the classification.

**Table 5** Classifier parameters

|  |  |
| --- | --- |
| Classifier | Paramaters |
| Decision tree | Maximum depth = 12, Use surrogates = false, Truncate pruned tree = true, Minimum sample count = 34, Priors = [0.33 0.33 0.33] |
| Random forest | Maximum number of trees = 5, Size of feature-set = 4, Maximum tree depth = 10, Forest accuracy = 0.025, Priors = [0.2 0.4 0.2] |
| kNN | K = 5, Priors = [0.33 0.33 0.33] |
| SVM | SVM type = C Support vector classification, Kernel = RBF, Kernel width = 25, C = 1, Priors = [0.33 0.33 0.33] |
| Bayes normal | Priors = [0.33 0.33 0.33] |

## Validation

The per-pixel performance of the candidate classifiers on the selected features was evaluated with the labeled pixel data. To avoid biased estimates of performance, ten-fold cross validation was used for classifier evaluation. The canopy-cover performance of the classifiers was tested on the in situ canopy-cover data. After applying the classifiers and morphological operations to the relevant images, canopy-cover estimates were extracted by evaluating the fractional portion of spekboom inside the areas of the field site polygons. These estimates were compared to the in situ canopy-cover data.

# Results

## Feature Selection

The dendrogram showing the clustering of our feature set, is plotted in Fig. 6. The red line shows the dissimilarity threshold value at which the feature clusters were extracted. This value was selected on the basis of being a relatively stable point in the hierarchy and being a point where the correlation among features is strong.



Fig. 6 Clustering of correlated features

Table 6 lists the clusters ordered by their importance, along with their component features.

**Table 6** Ranked clusters

|  |  |  |
| --- | --- | --- |
| Rank | Importance (%) | Features |
| 1 | 68.27 | rN, nirN, NDVI, RVI, tc2, pc2, nc1, MeanRVI, MedianRVI, MeanNDVI, MedianNDVI |
| 2 | 61.38 | R, G, B, NIR, tc1, pc1, MeanPc1, MedianPc1 |
| 3 | 60.41 | EntropyPc1 |
| 4 | 55.23 | gN, MeanGn, MedianGn |
| 5 | 54.52 | bN |
| 6 | 53.57 | nc2, nc4 |
| 7 | 50.57 | tc4, nc3 |
| 8 | 49.34 | pc4 |
| 9 | 47.93 | EntropyRVI, StdRVI, EntropyNDVI, StdNDVI |
| 10 | 43.96 | StdPc1 |
| 11 | 43.62 | EntropyGn, StdGn |
| 12 | 42.65 | tc3, pc3 |
| 13 | 41.29 | SkewnessRVI, SkewnessNDVI |
| 14 | 35.27 | SkewnessGn |
| 15 | 35.19 | KurtosisRVI, KurtosisNDVI |
| 16 | 35.03 | SkewnessPc1 |
| 17 | 34.86 | KurtosisGn |
| 18 | 33.86 | KurtosisPc1 |

The NDVI, pc1, EntropyPc1, gN, bN and nc2 features were selected from the top six clusters.

## Classification and Canopy-Cover Estimation

Table 7 compares the performance of the candidate classifiers. The table results are sorted according to the mean absolute canopy-cover error (MAE) in the last column. Of the performance measures in the table, this is the only one evaluated against the in situ canopy-cover data; the rest were evaluated against the labeled pixel data using cross-validation. Three- and two-class errors are reported as the class prior weighted errors i.e. the mean of the errors of omission. Cohen’s Kappa and consumer’s and producer’s accuracies are given for the two-class case. Standard errors are given for all cross-validated performance measures evaluated on the labelled pixel data. The standard deviation of absolute errors (SAE) gives an indication of the variability in the canopy cover performance. Times taken for each classifier to process a single 12000 pixel × 8000 pixel image are listed in Table 8.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | a | a |  |  | a |  |
|  | 9.49 ± 0.47 | 3.59 ± 0.31 | 95.26 ± 0.44 / 98.05 ± 0.37 | 98.32 ± 0.32 / 94.50 ± 0.53 | 0.930 ± 0.006 |  |
|  | 9.15 ± 0.95 | 2.69 ± 0.23 | 97.24 ± 0.25 / 97.45 ± 0.30 | 97.74 ± 0.27 / 96.89 ± 0.28 | 0.947 ± 0.005 |  |
|  | 10.28 ± 0.41 | 1.70 ± 0.13 | 98.95 ± 0.18 / 97.52 ± 0.25 | 97.05 ± 0.42 / 98.84 ± 0.20 | 0.965 ± 0.002 |  |
|  | 9.98 ± 1.51 | 2.47 ± 1.95 | 98.90 ± 0.21 / 96.08 ± 3.85 | 96.27 ± 3.90 / 98.80 ± 0.22 | 0.949 ± 0.041 |  |
|  | 16.31 ± 0.86 | 9.03 ± 0.50 | 86.95 ± 0.62 / 98.09 ± 0.28 | 98.55 ± 0.21 / 83.38 ± 0.88 | 0.826 ± 0.010 |  |

± = standard error of cross validated performance measure,

Table 8 Classifier computation times

|  |  |
| --- | --- |
| Classifier | Time (secs)a |
| Decision tree | 47 |
| Random forest | 138 |
| kNN | 2067 |
| SVM | 788 |
| Bayes normal | 61 |

a computation time per 12000 pixel ×8000 pixel image

The decision tree three-class and two-class confusion matrices and performances, obtained from the labeled pixel data, are given in Table 8 and Table 9 respectively. The three-class confusion matrix shows that the tree class overlaps with both the spekboom and background classes, but that the overlap is larger with the background class. Table 10 shows the canopy-cover estimates obtained from the post-processed decision tree output for each of the in situ canopy-cover sites. The mean of the absolute canopy-cover error is 5.85%, with a standard deviation of 4.65%.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
|  | 24776 | 330 | 2154 | 27260 | 90.89 ± 0.49 |
|  | 297 | 25762 | 1201 | 27260 | 94.50 ± 0.53 |
|  | 282 | 183 | 2892 | 3357 | 86.15 ± 1.33 |
|  | 25355 | 26275 | 6247 | 57877 |  |
|  | 97.72 ± 0.29 | 98.05 ± 0.37 | 46.34 ± 1.50 |  |  |
|  | 0.866 ± 0.007 |  |  |  |  |
|  | 9.49 ± 0.47 |  |  |  |  |

± = standard error of cross validated performance measure,

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  | 30104 | 513 | 30617 | 98.32 ± 0.32 |
|  | 1498 | 25762 | 27260 | 94.50 ± 0.53 |
|  | 31602 | 26275 | 57877 |  |
|  | 95.26 ± 0.44 | 98.05 ± 0.37 |  |  |
|  | 0.930 ± 0.006 |  |  |  |
|  | 3.59 ± 0.31 |  |  |  |

± = standard error of cross validated performance measure,

**Table 10** Decision tree canopy-cover estimates

|  |  |  |  |
| --- | --- | --- | --- |
| Area | No. | Ground Truth (%) | Classifier (%) |
| Groenfontein | 1 | 0.00 | 0.07 |
|  | 2 | 4.00 | 0.47 |
|  | 3 | 10.00 | 8.21 |
|  | 4 | 25.00 | 17.44 |
| Matjiesvlei | 1a | 6.00 | 7.21 |
|  | 1b | 22.50 | 31.37 |
|  | 2 | 70.00 | 67.38 |
|  | 3 | 85.00 | 73.12 |
|  | 4 | 65.00 | 70.34 |
|  | 5 | 37.50 | 35.95 |
|  | 6 | 17.50 | 12.01 |
|  | 7 | 15.00 | 25.74 |
|  | 8 | 2.00 | 5.42 |
| Rooiberg | 1 | 20.00 | 6.03 |
|  | 2 | 11.00 | 1.03 |
|  | 3 | 0.00 | 0.00 |
| Grootkop | 1 | 22.50 | 8.05 |
|  | 2 | 0.50 | 0.22 |
|  | 3 | 42.50 | 34.38 |
|  | 4 | 77.50 | 71.27 |
| **MAE (SAE)a** | | | **5.85% (4.65%)** |

aMAE = Mean of absolute canopy-cover errors (%), SAE = Standard deviation of absolute canopy-cover errors (%)

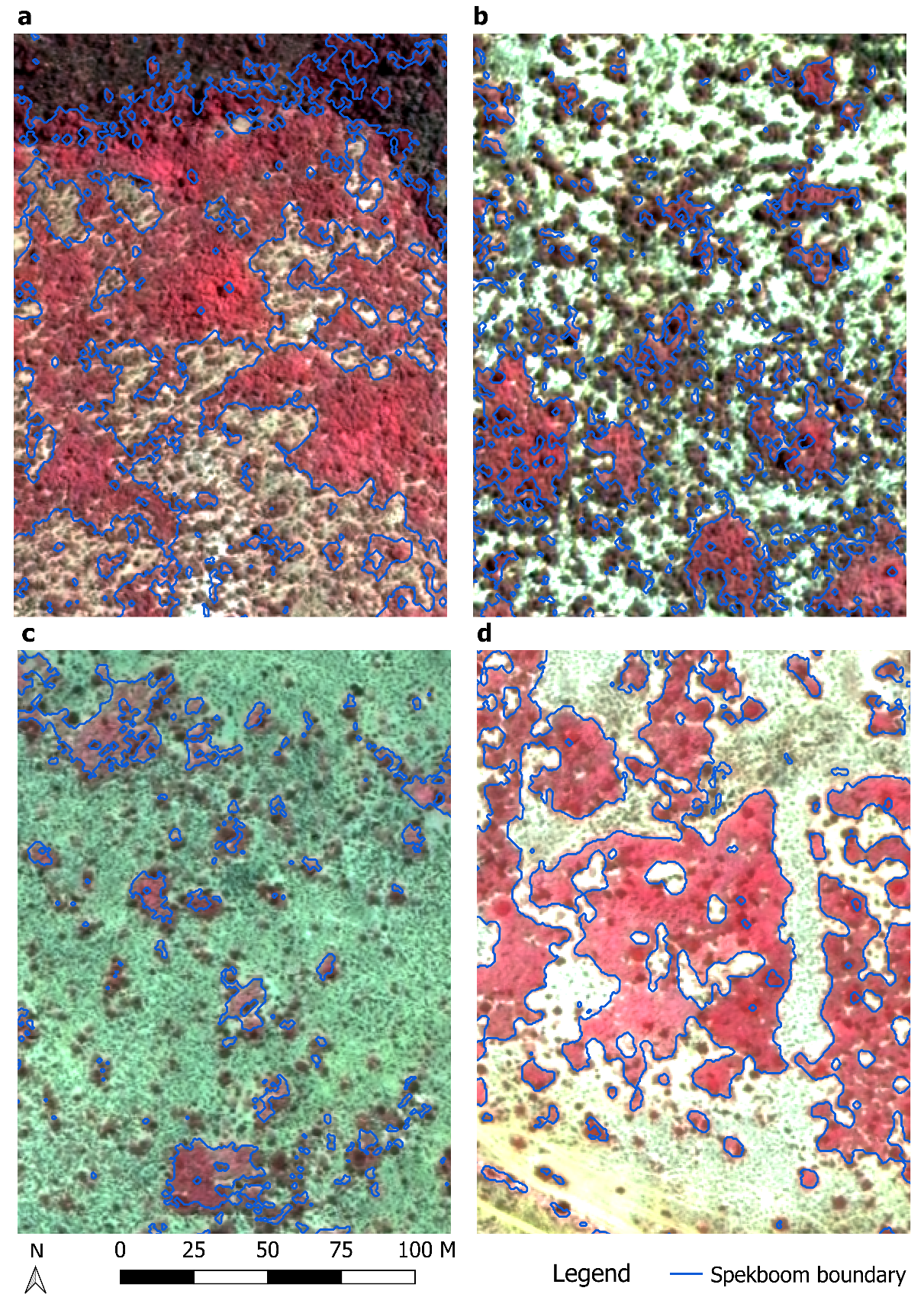
The decision tree classifier was applied to the image mosaic of the study area to produce a spekboom canopy-cover map that was morphologically post-processed. Fig. 7 shows close-up examples of the resulting canopy-cover boundary, overlaid on a color-infrared (CIR) rendering of the aerial imagery, for each of the canopy-cover ground truth areas (as described in Table 1). 

Fig. 7 Example canopy cover maps showing Spekboom boundaries (a) Matjiesvlei (Habitat: arid thicket with spekboom), b) Groenfontein (Habitat: valley thicket with spekboom); (c) Rooiberg (Habitat: arid thicket with spekboom and fynbos mosaic); and (d) Grootkop (Habitat: arid thicket with spekboom and succulent Karoo mosaic)

# Discussion

## Feature Selection

Table 6 reveals a number of interesting properties of the features. First, it is clear that there is significant redundancy among the features. The correlation between the R, G, B and NIR bands is strong (>0.7), likely due to strong coupling with intensity. The bands are consequently all grouped into a single cluster. While the definitions of the nirN, NDVI and RVI features are quite different, they are all describing the same spectral property of vegetation, namely high absorption in the red band and high reflectance in the NIR band. This is confirmed by their collection in the same cluster.

EntropyPc1 is ranked highly (third) in its own cluster, which supports the hypothesis that texture is an important property for mapping vegetation in VHR imagery. It is, however, the only texture feature in the best eight clusters. At the 0.5 m image resolution, texture will be descriptive of bush-clumps more than individual spekboom plants. The bush-clumps vary significantly in their composition and character with variation in habitat and level of degradation. We believe that the paucity of texture features in informative clusters is likely due to bush-clump and shadow variations.

The importance of bN was unexpected. The blue channel is particularly susceptible to haze effects and intuitively should not hold much discriminating power for vegetation. Inspecting bN images shows an inversion of the topography shading seen in other channels. Sunlit northern slopes are dimmer and shaded southern slopes brighter. This occurs because the blue light in the shaded areas, which scatters more readily, is the dominant band of illumination. The contribution of bN is not fully understood but we believe its value lies in this property and that it helps to distinguish shaded vegetation from genuinely dark vegetation. In their tree mapping study, Key et al.58 also found the blue band to be valuable due to its insensitivity to shadowing issues.

The gN feature, its mean and its median form their own cluster. The mean sliding window feature, median sliding window feature and source feature operated on by those sliding windows are strongly correlated, as is expected.

The NDVI, pc1, EntropyPc1, gN, bN and nc2 features were selected from the top six clusters. Selection of sliding window features was avoided where possible as they are computationally more demanding than the per-pixel features. NDVI was selected from the first cluster simply because it is popular and easy to interpret. In the second cluster, pc1 was chosen as it is the first principal component of the raw bands and should therefore be more informative than any one of them in isolation. There is only one sliding window feature, EntropyPc1, in our final selection.

## Classification and Canopy-Cover Estimation

With the exception of the Bayes normal classifier, the classifiers’ performance was remarkably good. The performances of the kNN and decision tree classifiers are as good as or better than the more complex SVM and random forest classifiers (see Table 7). The excellent performance of a diverse group of classifiers suggests that an informative feature set was selected. The notably poorer performance of the Bayes normal classifier implies that the classes are not normally distributed. The three-class errors are larger than the two-class errors due the tree class overlapping substantially with the background class. Errors due to tree samples being assigned to the background class, and vice versa, are negated when the tree class is lumped into the background class. The classifier performance variability on the labeled pixel data was low in general, as evidenced by the standard errors in Table 7, with the SVM producing the highest variability of the tested classifiers.

Of the performance measures in Table 7, the MAE is considered the most important for classifier comparison as it has the most direct relationship with actual canopy-cover mapping accuracy over the study area. Taking the MAE and image ground truth performance into account, the decision tree was selected as the final classifier. It has the best canopy-cover performance and is the fastest option (see Table 8). While it is one of the poorer performers on the labeled pixel data, it is still very accurate when applied to this data.

The classifier performed well in the Groenfontein, Matjiesvlei and Grootkop areas, but underestimated canopy cover in all the Rooiberg sites. As a result of the sandstone/quartzite geology of the area, the spekboom plants at Rooiberg are smaller and have a canopy that is less dense than those in other sites. We believe this partially explains the canopy-cover underestimation in this area.

A visual inspection of the canopy-cover map revealed some spatial variation over the study area. Fig. 7 shows close-up canopy-cover map examples for each of the canopy-cover ground truth areas (as described in Table 1). Arid areas, such as Rooiberg, seem more prone to underestimation, probably due to spectral mixing occurring with bare ground around the canopy borders and also due to the smaller and less dense stands occurring in these areas. Conversely, there tends to be a slight overestimation in more densely vegetated areas, likely the result of confusion due to spectral mixing with other green vegetation. In general, however, the canopy-cover map of the study area appears accurate.

This study is one of few examples of vegetation mapping using VHR imagery over a large area.12,14. brationro conflicts of interest to declare While the mapping accuracies achieved compare well with related studies 18,19,23,26,59, there are possible avenues for improvement. Ancillary information such as elevation, slope aspect or habitat could be incorporated into the classifier, similarly to Thompson et al.Thompson et al.6 and Su60. This could be done by including it either as a feature or by designing separate classifiers for different ranges or categories of the ancillary variable.

# Conclusions

Accurate spekboom canopy-cover estimates were obtained across the study area using a per-pixel classification approach. Homogenization to surface reflectance by calibration with satellite data provided radiometric consistency and allowed application of a single classification algorithm over an extended area. Six features, consisting of a combination of spectral, textural and vegetation index type measures, were selected using a feature clustering and ranking method. Out of a set of candidate classifiers, a decision tree produced the best canopy-cover accuracy, and was subsequently used to produce a map of the study area. A MAE of 5.86% over 20 ground truth sites was achieved.

While some variation in the canopy-cover accuracy was observed over different habitats, the classifier’s general performance was consistent over the study area. By incorporating ground truth from new areas, the techniques used to produce this map could be applied to the rest of the thicket biome. The availability of a spekboom canopy-cover mapping technique will be a valuable starting point for developing measures of other environmental variables such as biomass and biodiversity.12,61,62

## Table of Figures

Fig. 1 Little Karoo study area 6

Fig. 3 Study area spekboom habitats and field ground truth sites 8

Fig. 4 Matjiesvlei2 canopy-cover ground truth site 9

Fig. 5 Example image class labels 11

Fig. 2 MODIS and DMC RSR’s 12

Fig. 6 Clustering of correlated features 21

Fig. 7 Groenfontein classification (Habitat: valley thicket with spekboom) 26

Fig. 8 Matjiesvlei classification (Habitat: arid thicket with spekboom) 26

Fig. 9 Rooiberg classification (Habitat: arid thicket with spekboom and fynbos mosaic) 27

Fig. 10 Grootkop classification (Habitat: arid thicket with spekboom and succulent Karoo mosaic) 27

## Disclosure

There are no conflicts of interest to declare.

## Acknowledgments

We would like to thank Adrian Roos and Intergraph South Africa for providing a license for Intergraph PPS, Bernard Jacobs of Geospace International for facilitating insight into the NGI image processing workflow and in obtaining DMC relative spectral response data, Theo Pauw and Garth Stephenson of CGA for assistance with computing and software resources, and Julie Verhulp and NGI for provision of the aerial imagery. The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF. This work was supported by funding from the Department of Environmental Affairs via the Working for Natural Resources Program.

## References

1. J. Vlok, R. M. Cowling, and T. Wolf, “A vegetation map for the Little Karoo,” Unpublished maps and report for a SKEP project supported by CEPF grant no 1064410304 (2005).

2. 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, Argentina (2009).

3. 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].

4. 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).

5. 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).

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. G. van Luijk et al., “Hydrological implications of desertification: Degradation of South African semi-arid subtropical thicket,” J. Arid Environ. **91**, 14–21, Elsevier Ltd (2013) [doi:10.1016/j.jaridenv.2012.10.022].

8. 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).

9. M. L. Vyver et al., “Active restoration of woody canopy dominants in degraded South African semi-arid thicket is neither ecologically nor economically feasible,” Appl. Veg. Sci. **15**(1), 26–34 (2012) [doi:10.1111/j.1654-109X.2011.01162.x].

10. H. Adie and R. I. Yeaton, “Regeneration dynamics in arid subtropical thicket, South Africa,” South African J. Bot. **88**, 80–85, South African Association of Botanists (2013) [doi:10.1016/j.sajb.2013.05.010].

11. A. J. Mills and R. M. Cowling, “Below-ground carbon stocks in intact and transformed subtropical thicket landscapes in semi-arid South Africa,” J. Arid Environ. **74**(1), 93–100, Elsevier Ltd (2010) [doi:10.1016/j.jaridenv.2009.07.002].

12. 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).

13. M. J. Powell, “Restoration of degraded subtropical thickets in the Baviaanskloof Megareserve, South Africa,” Rhodes University (2009).

14. 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].

15. P. Curran et al., “Mapping restoration opportunity for collaborating with land managers in a carbon credit-funded restoration program in the Makana municipality, Eastern Cape, South Africa,” Restor. Ecol. **20**(1), 56–64 (2012) [doi:10.1111/j.1526-100X.2010.00746.x].

16. A. J. Mills et al., *Investing in sustainability. Restoring degraded thicket, creating jobs, capturing carbon and earning green credit.*, 1234, in Published by Climate Action Partnership, Cape Town, and Wilderness Foundation, Port Elizabeth, Climate Action Partnership, Cape Town and Wilderness Foundation, Port Elizabeth. (2010).

17. S. Basu et al., “A Semiautomated Probabilistic Framework for Tree-Cover Delineation From 1-m NAIP Imagery Using a High-Performance Computing Architecture,” IEEE Trans. Geosci. Remote Sens. **53**(10), 5690–5708 (2015) [doi:10.1109/TGRS.2015.2428197].

18. A. Ghosh and P. K. Joshi, “A comparison of selected classification algorithms for mapping bamboo patches in lower Gangetic plains using very high resolution WorldView 2 imagery,” Int. J. Appl. Earth Obs. Geoinf. **26**, 298–311, Elsevier B.V. (2014) [doi:10.1016/j.jag.2013.08.011].

19. K. Johansen et al., “Application of high spatial resolution satellite imagery for riparian and forest ecosystem classification,” Remote Sens. Environ. **110**(1), 29–44 (2007) [doi:10.1016/j.rse.2007.02.014].

20. S. Kollár, Z. Vekerdy, and B. Márkus, “Aerial image classification for the mapping of riparian vegetation habitats,” Acta Silv. Lignaria Hungarica **9**(1), 119–133 (2013) [doi:10.2478/aslh-2013-0010].

21. Y. T. Mustafa and H. N. Habeeb, “Object based technique for delineating and mapping 15 tree species using VHR WorldView-2 imagery,” in Proc. SPIE **9239**, p. 92390G–92390G–13 (2014) [doi:10.1117/12.2067280].

22. Z.-T. Ouyang et al., “A comparison of pixel-based and object-oriented approaches to VHR imagery for mapping saltmarsh plants,” Ecol. Inform. **6**(2), 136–146, Elsevier B.V. (2011) [doi:10.1016/j.ecoinf.2011.01.002].

23. A. Baraldi et al., “Automatic spectral-rule-based preliminary classification of radiometrically calibrated SPOT-4/-5/IRS, AVHRR/MSG, AATSR, IKONOS/QuickBird/OrbView/GeoEye, and DMC/SPOT-1/-2 imagery—Part I: System design and implementation,” IEEE Trans. Geosci. Remote Sens. **48**(3), 1299–1325 (2010).

24. J. Boyden et al., “Utility of VHR remote sensing data for landscape scale assessment of the environmental weed para grass [UROCHLOA MUTICA, (FORSSK), NGUYEN] on a tropical floodplain,” in 28th Asian Conference on Remote Sensing 2007, ACRS 2007 **1**, pp. 551–556 (2007).

25. A. I. de Castro et al., “Airborne multi-spectral imagery for mapping cruciferous weeds in cereal and legume crops,” Precis. Agric. **13**(3), 302–321 (2012) [doi:10.1007/s11119-011-9247-0].

26. H. Mehner et al., “Remote sensing of upland vegetation: the potential of high spatial resolution satellite sensors,” Glob. Ecol. Biogeogr. **13**(4), 359–369 (2004) [doi:10.1111/j.1466-822X.2004.00096.x].

27. Trimble, *eCognition Developer user guide* (2016).

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

29. 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).

30. L. Tolosi and T. Lengauer, “Classification with correlated features: unreliability of feature ranking and solutions.,” Bioinformatics **27**(14), 1986–1994 (2011) [doi:10.1093/bioinformatics/btr300].

31. Z. Mingguo, C. Qianguo, and Q. Mingzhou, “The Effect of Prior Probabilities in the Maximum Likelihood Classification on Individual Classes,” Photogramm. Eng. Remote Sens. **75**(9), 1109–1117, IEEE, Seoul, South Korea (2009) [doi:10.14358/PERS.75.9.1109].

32. ESA, “Sentinel-2 User Handbook” (2015).

33. G. Schmidt et al., “Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) Algorithm Description” (2012) [doi:No. 2013-1057].

34. GDAL Development Team, “Geospatial Data Abstraction Library,” Open Source Geospatial Foundation, 2014, <http://www.gdal.org/>.

35. G. Bradski, “The OpenCV library,” Dr. Dobb’s J. Softw. Tools **25**(120), 122–125 (2000).

36. Z. Li et al., “Evaluation of spectral and texture features for object-based vegetation species classification using support vector machines,” in ISPRS TC VII Symposium – 100 Years ISPRS **XXXVIII**, W. Wagner and B. Székely, Eds., pp. 122–127, IAPRS, Vienna, Austria (2010).

37. 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].

38. M. Singh, Y. Malhi, and S. Bhagwat, “Biomass estimation of mixed forest landscape using a Fourier transform texture-based approach on very-high-resolution optical satellite imagery,” Int. J. Remote Sens. **35**(9), 3331–3349 (2014) [doi:10.1080/01431161.2014.903441].

39. 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).

40. D. M. Gates, *Biophysical Ecology*, Springer, New York (1980) [doi:10.1007/978-1-4612-6024-0].

41. R. B. Myneni et al., “The interpretation of spectral vegetation indexes,” IEEE Trans. Geosci. Remote Sens. **33**(2), 481–486 (1995) [doi:10.1109/36.377948].

42. R. J. Kauth and G. S. Thomas, “The tasselled cap -- a graphic description of the spectral-temporal development of agricultural crops as seen by LANDSAT,” in Symposium on Machine Processing of Remotely Sensed Data, p. 4B41-4B51, IEEE, Purdue University of West Lafayette, Indiana, USA (1976).

43. Intergraph, “Digital mapping camera system,” 2008, <http://www.geospace.co.za/pdf/DMC Brochure.pdf>.

44. B. Sahu and D. Mishra, “A novel approach for selecting informative genes from gene expression data using signal-to-noise ratio and t-statistics,” in 2011 2nd International Conference on Computer and Communication Technology (ICCCT-2011), S. Mohanty, Ed., pp. 5–10, IEEE, Allahabad, India (2011) [doi:10.1109/ICCCT.2011.6075207].

45. C. Strobl et al., “Conditional variable importance for random forests.,” BMC Bioinformatics **9**, 307 (2008) [doi:10.1186/1471-2105-9-307].

46. M. Yousef et al., “Recursive cluster elimination (RCE) for classification and feature selection from gene expression data.,” BMC Bioinformatics **8**(144) (2007) [doi:10.1186/1471-2105-8-144].

47. D. Harris and A. Van Niekerk, “Feature clustering and ranking for selecting stable features from high dimensional remotely sensed data,” Int. J. Remote Sens., 1–16, Taylor & Francis (2018) [doi:10.1080/01431161.2018.1500730].

48. G. J. Szekely and M. L. Rizzo, “Hierarchical clustering via joint between-within distances: extending Ward’s minimum variance method,” J. Classif. **22**(2), 151–183 (2005) [doi:10.1007/s00357-005-0012-9].

49. D. J. Hand and K. Yu, “Idiot’s Bayes - Not So Stupid After All?,” Int. Statisitical Rev. **69**(3), 385–398 (2001).

50. L. Breiman et al., *Classification and regression trees*, Wadsworth, Calif. (1984).

51. L. Breiman, “Bagging predictors,” Mach. Learn. **24**(2), 123–140 (1996).

52. L. Breiman, “Random Forests,” Mach. Learn. **45**(1), 5–32 (2001) [doi:10.1023/A:1010933404324].

53. J. Amorós López et al., “Land cover classification of VHR airborne images for citrus grove identification,” ISPRS J. Photogramm. Remote Sens. **66**(1), 115–123, Elsevier B.V. (2011) [doi:10.1016/j.isprsjprs.2010.09.008].

54. C. J. C. Burges, “A tutorial on support vector machines for pattern recognition,” Data Min. Knowl. Discov. **2**(2), 121–167 (1998) [doi:10.1023/A:1009715923555].

55. R. P. W. Duin and D. M. J. Tax, “Statistical Pattern Recognition,” in Handbook of Pattern Recognition and Computer Vision, 3rd ed., C. Chen and P. Wang, Eds., pp. 1–21, World Scientific, Singapore (2005) [doi:10.1142/9789812775320\_0001].

56. OpenCV Development Team, “OpenCV documentation,” Open Source Computer Vision Library, 2014, <http://docs.opencv.org/>.

57. J. Serra and P. Soille, Eds., “Mathematical morphology and its applications to image processing,” in 2nd International Symposium on Mathematical Morphology (ISMM’94), p. 383, Kluwer Academic Publishers (1994).

58. T. Key et al., “A comparison of multispectral and multitemporal information in high spatial resolution imagery for classification of individual tree species in a temperate hardwood forest,” Remote Sens. Environ. **75**(1), 100–112, Elsevier (2001).

59. B. W. Heumann, “Satellite remote sensing of mangrove forests: Recent advances and future opportunities,” Prog. Phys. Geogr. **35**(1), 87–108 (2011) [doi:10.1177/0309133310385371].

60. L. Su, “Optimizing support vector machine learning for semi-arid vegetation mapping by using clustering analysis,” ISPRS J. Photogramm. Remote Sens. **64**(4), 407–413 (2009) [doi:10.1016/j.isprsjprs.2009.02.002].

61. H. Suganuma et al., “Stand biomass estimation method by canopy coverage for application to remote sensing in an arid area of Western Australia,” For. Ecol. Manage. **222**(1–3), 75–87 (2006) [doi:10.1016/j.foreco.2005.10.014].

62. J. A. Ludwig, J. F. Reynolds, and P. D. Whitson, “Size-biomass relationships of several Chihuahuan desert shrubs,” Am. Midl. Nat. **94**(2), 451–461 (1975).