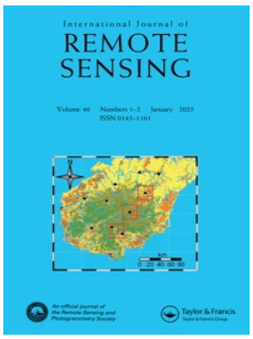

International Journal of Remote Sensing

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tres20

Determining the value of different wavelength ranges of non-imaging hyperspectral reflectance to estimate carotenoid content using the PROSPECT-5 model

Bongokuhle Sibiya, Moses Cho, Onisimo Mutanga, John Odindi, Cecilia Masemola & Wessel Bonnet

To cite this article: Bongokuhle Sibiya, Moses Cho, Onisimo Mutanga, John Odindi, Cecilia Masemola & Wessel Bonnet (2025) Determining the value of different wavelength ranges of non-imaging hyperspectral reflectance to estimate carotenoid content using the PROSPECT-5 model, International Journal of Remote Sensing, 46:4, 1696-1719, DOI: 10.1080/01431161.2024.2439081

To link to this article: https://doi.org/10.1080/01431161.2024.2439081

Published online: 28 Dec 2024.

INTERNATIONAL JOURNAL OF REMOTE SENSING
2025, VOL. 46, NO. 4, 1696–1719
https://doi.org/10.1080/01431161.2024.2439081

Taylor & Francis
Taylor & Francis Group

Determining the value of different wavelength ranges of non-imaging hyperspectral reflectance to estimate carotenoid content using the PROSPECT-5 model

Bongokuhle Sibiya[a,b], Moses Cho[b], Onisimo Mutanga[a], John Odindi[a], Cecilia Masemola[b] and Wessel Bonnet[b]

[a]Discipline of Geography, School of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal, Pietermaritzburg, South Africa; [b]Precision Agriculture Research Group, Advanced Agriculture and Food, Council for Scientific and Industrial Research (CSIR), Pretoria, South Africa

ABSTRACT

Carotenoids are important plant attributes offering valuable insights into the physiological condition of vegetation and serve as essential indicators for early identification of plant stress. Generally, carotenoids can be extracted using radiative transfer model (RTM) remote sensing techniques like PROSPECT that utilize the entire spectral domain (400–2500 nm) to retrieve carotenoid information. However, such inversions suffer from ill-posed due to model uncertainties. Literature suggests that selecting appropriate bands improves the RTM inversion. Hence, this study proposed a wavelength selection approach using various regions in the visible portion of the electromagnetic spectrum and bands selected by the random forest algorithm to estimate carotenoids using the PROSPECT-5 model. This study utilized three distinct datasets – savanna, tropical forest, and a combination of two. The green spectral region demonstrated the strongest performance in the tropical forest dataset ($R^2 = 0.90$, RMSE = 0.71) than the savanna ($R^2 = 0.70$, RMSE = 1.19) and combined ($R^2 = 0.72$, RMSE = 1.11) datasets, respectively. The bands (green, yellow, and red-edge) selected by the random forest model produced the highest accuracy in the savanna dataset ($R^2 = 0.84$, RMSE = 0.99), followed by combined ($R^2 = 0.80$, RMSE = 1.20) and tropical forest ($R^2 = 0.78$, RMSE = 1.33), respectively. Lastly, the visible region demonstrated strong performance in the tropical forest ($R^2 = 0.84$, RMSE = 0.85), followed by combined datasets ($R^2 = 0.72$, RMSE = 1.15) and savanna ($R^2 = 0.68$, RMSE = 1.27), respectively. The findings suggest that carotenoid retrieval should be limited to the visible portions of the spectrum as it exhibited strong performance in estimating carotenoid content across the savanna, tropical forest, and combined datasets.

ARTICLE HISTORY
Received 31 December 2023
Accepted 18 November 2024

KEYWORDS
Analytical Spectral Device (ASD); carotenoids; PROSPECT-5; and Radiative Transfer Model (RTM)

1. Introduction

Savannas and indigenous forests are distinct ecosystems with unique vegetation, biodiversity, and ecological functions. The savanna, which consists of herbaceous layer of grass

CONTACT Bongokuhle Sibiya bongokuhlesibiya12@gmail.com Precision Agriculture Research Group, Advanced Agriculture and Food, Council for Scientific and Industrial Research (CSIR), Pretoria 0001, South Africa

species and open layer of trees, is the largest biome in Southern Africa, covering approximately 46% of the landscape. In South Africa, it covers one-third (399,600 km^2) of the county's landmass (Minasny et al. 2017). On the other hand, indigenous forests, accounting 0.56% of the country's landscape, are characterized by contiguous tree crowns within small patches, typically less than 1 km^2 (Cho, Malahlela, and Ramoelo 2015; Lawes, Macfarlane, and Eeley 2004). These ecosystems are crucial for carbon sequestration and climate change mitigation (Odebiri et al. 2023). They also support biodiversity, offer habitats for wildlife, and provide a range of socio-economic benefits and services, including, food, wood, and medicine (Madonsela et al. 2017). However, these ecosystems face significant threats. For instance, over the last decades, approximately 20% of South Africa's biomes have undergone conversion to pasture, farmland, and urban developments (Odebiri 2022). Consequently, intensive management initiatives are necessary to preserve the savanna and the remaining indigenous forest patches to prevent further loss.

To prevent losses and restore plant biodiversity, quantifying and modelling its functional diversity, measured using plant attributes is paramount (J. T. Zhang, Fan, and Li 2012). Carotenoids are essential plant attributes valuable for ecosystem functioning and plant physiology status, which can provide valuable information to effectively assess how anthropogenic and natural factors such as climate change affect plant adaptation and functioning (Féret et al. 2017; Y. Zhang et al. 2021). Generally, carotenoids are regarded as reliable indicators of vegetation's physiological status (Huang et al. 2018). According to Zhou et al. (2017), a decrease in chlorophyll at various stages of a plant's growth cycle indicates that it is experiencing environmental stress. On the other hand, alterations in carotenoid concentrations indicate vegetation's physiological conditions. Literature also reveals that carotenoid content changes when plants are exposed to intense sunlight, high temperatures, low nitrogen, and onset of leaf senescence (Huang et al. 2018). Therefore, quantifying carotenoid content is essential for early detection of vegetation stress and detecting the senescence cycle, as well as elucidating photoprotection and light adaptation mechanisms.

In the past decades, remote sensing (RS), with the benefits of being non-destructive, has been one of the most cost-effective approaches for retrieving plant attributes such as carotenoids (Mutanga and Skidmore 2004; Sibiya, Lottering, and Odindi 2021; Y. Zhang et al. 2021). The approach has also shown great potential to quickly estimate carotenoids over large spatial extents (Asner et al. 2002; Sonobe and Hirono 2023).

Previous studies have demonstrated progress in estimating carotenoid content using remotely sensed data, both at the leaf and canopy levels (Huang et al. 2018). However, much of this research has primarily focused on the development of spectral indices for such estimation (Blackburn 1998; Chappelle, Kim, and McMurtrey 1992; Datt 1998; Gitelson et al. 2002; Gitelson, Keydan, and Merzlyak 2006; Hernández-Clemente, Navarro-Cerrillo, and Zarco-Tejada 2012). Spectral indices or empirical models generally require calibration for each specific dataset, as it is influenced by the characteristics of the given database. This constraint affects the reliability and adaptability of the method in different settings (Li et al. 2018; Zhou et al. 2017).

On the other hand, physical models such as the PROSPECT (Feret et al. 2008) are rooted in understanding physical processes governing the interaction between photons and plant biochemical characteristics, which make them suitable for retrieving plant attributes such as carotenoids, as they are more robust and transferable than empirical models (Cho,

Ramoelo, and Skidmore 2014; Wan et al. 2022). The PROSPECT-5 radiative transfer model, developed by Feret et al. in 2008, is one of the most used versions within the PROSPECT family of models. This model incorporates two types of pigments related to photosynthesis in vegetation: chlorophylls and carotenoids. The leaf samples used to calibrate the PROSPECT-5 model are primarily green in colour. As a result, this model is effective at simulating spectra and estimating the biochemical and biophysical properties of photosynthetically active leaves (Lu, Proctor, and He 2021; Proctor, Lu, and He 2017). Given that the primary aim of this study was to estimate carotenoids, the PROSPECT-5 model provides a robust framework as it can accurately simulate the reflectance characteristics of vegetation by incorporating the optical properties of these key pigments (Lu, Proctor, and He 2021). Specifically, the PROSPECT-5 model provides a valuable tool for estimating carotenoid content from leaf-level spectral data, yielding insights into the health and functioning of the studied vegetation, crucial for a wide range of remote sensing and ecological applications.

Generally, the PROSPECT inversion is employed to simulate wavelength across the entire optical domain (400 nm–2500 nm), which is frequently unavailable and demands high computational processing resources (He et al. 2023). However, considering the entire spectral range, including less informative wavelengths, the inversion process may introduce unnecessary noise or inaccuracies. To address these issues and enhance the accuracy of retrieving leaf biochemical traits, a carefully selected subset of wavelengths or bands most sensitive to the desired properties for the inversion process is necessary (Song et al. 2021; Song et al. 2011). According to the literature, selecting suitable wavelength improves the inversion and reduces biased variable estimation caused by measurement and model uncertainty (Sun et al. 2019; J. Yang et al. 2021). For instance, literature reveals that the unique absorption properties of carotenoids (Car) in the visible portion of the spectrum allow for quantifying carotenoids content using remote sensing approaches. Thus, this study utilized the PROSPECT-5 model to investigate band selection method to estimate carotenoids. The choice of PROSPECT-5 was driven by the dataset's summer-only collection, which lacks seasonal variations. This model's focus on simulating the optical properties of mature, photosynthetic leaves makes it particularly suitable, offering greater accuracy compared to models like PROSPECT-D and PROSPECT-PRO that account for the entire leaf lifecycle (Lu, Proctor, and He 2021). Hence, this study sought to investigate band selection method in the visible region of the spectra using PROSPECT-5 model to estimate carotenoids in the savanna and tropical forest using Artificial Neural Network (ANN) inversion.

The adoption of an Artificial Neural Network (ANN) as a hybrid approach in the PROSPECT-5 model offers several benefits compared to other machine learning algorithms. First, ANNs are recognized for their superior ability to capture complex non-linear relationships within the data, a crucial advantage when working with the intricate radiative transfer processes modelled by PROSPECT-5 (Verrelst et al. 2015, 2019). Unlike more rigid parametric models, the flexible data-driven nature of ANNs allows them to adapt and learn the underlying patterns in the spectral data, leading to more accurate estimations of leaf biochemical constituents like carotenoids.

In addition to the wavelength selection method in the visible region of the electromagnetic spectrum, the present study also investigated the performance of various machine-learning algorithms for feature selection. Random Forest was chosen for its

ability to handle nonlinear relationships and provide insights into feature importance (Mngadi, Odindi, and Mutanga 2021). XGBoost was selected due to its superior predictive performance and capacity to identify the most informative wavelengths (Sauer et al. 2022). Partial Least Square Regression (PLSR) was employed due to its effectiveness in dealing with high-dimensional data and identifying latent variables (C. Yang et al. 2023). Artificial Neural Network (ANN) was utilized for its powerful nonlinear modelling capabilities and potential to uncover complex spectral patterns (O. Odebiri, Mutanga, and Odindi 2022). The wavelengths identified through these feature selection techniques were subsequently incorporated into the PROSPECT-5 model to enhance the model's performance and accuracy.

Hence, this study's objectives were to i) perform feature selection using different algorithms such random forest, XGBoost, Partial Least Square Regression (PLSR), and Artificial Neural Network (ANN), ii) conduct the PROSPECT-5 inversion by employing the bands selected from the algorithm that exhibited the best performance, and iii) lastly utilize different regions within the visible spectrum (blue, green, yellow, red, and red-edge) during the PROSPECT-5 inversion process and evaluate the model's performance using three datasets: savanna, tropical forest, and a combined dataset.

2. Methods

2.1. Leaf spectral data

2.1.1. Location of the study plots

(a) **Kruger National Park (Savanna biome)**: The study site is situated within the savanna biome (latitude 24° 51′ 35.31″ to 24° 54′ 22.73″S, longitude 31° 20′ 37.88″E, and 31° 33′ 28.40″E). It is the largest biome in Southern Africa, covering approximately 46% of the region's landscape. It is characterized by herbaceous layer of various grass species and complemented by open tree canopies. Typically, the savanna undergoes a prolonged dry winter and a wet summer, with annual precipitation ranging from 235 to 1000 mm. The region of interest is primarily situated in the southern of the Kruger National Park and encompasses two study regions, which include Sabi Sands Wildtuin and Bushbuckridge Municipal District. The majority of Bushbuckridge is composed of granite soil types, while the Sabi Sands region still has gabbro intrusions (see Cho et al. 2012 and Naidoo et al. 2012 spatial distribution of the landscape).

(b) **Dukuduku indigenous coastal forest (Tropical Forest)**: Dukuduku forest is situated between Mtubatuba and St Lucia in the northern part of KwaZulu-Natal, South Africa (28° 25′ S, 32° 17′ E). The area covered approximately 6,000 hectares in the early 1950s; however, it dwindled due to significant deforestation. By 2011, its size had diminished to a mere 3,200 hectares, primarily due to conversion to farmlands and establishment of settlements (Ndlovu et al. 2011; Qabaqaba et al. 2023). The region receives consistent yearly rainfall, with an average annual precipitation of around 1,200 mm. The area was selected for investigation because it represents the largest remaining section of an indigenous forest (see Cho, Malahlela, and Ramoelo 2015 for the distribution of the study site).

2.2. Field survey

Similar sampling and methodological approach were employed at both the Dukuduku Tropical Forest and Savanna study sites. For the Dukuduku Tropical Forest, seven transect lines were randomly established using a pan-sharpened WorldView-2 satellite image. Each transect spanned a minimum distance of 1 km, and the tree crowns visible on the enhanced image were assigned numerical identifiers along the transects. The coordinates of these numbered tree crowns were then extracted and uploaded into a handheld Garmin Vista Cx etrex GPS device. In the Savanna landscape, the coordinates of the tree crowns were extracted from the satellite imagery using a network of roads and paths, focusing on a range of 0 to 200 m away from the road or path. These coordinates were also downloaded into the same handheld GPS device. During the survey, the pre-selected tree canopies were located and identified using a GPS device. Some tree species were found to be dominant in each landscape, as reported in previous studies (Cho et al. 2012; Cho, Malahlela, and Ramoelo 2015; Naidoo et al. 2012).

Field data was collected in May 2010 and August 2011 for the Savanna landscape and Dukuduku Tropical Forest, respectively, to record tree species and canopy gap. Leaf samples were collected using the detached branch technique targeting the predominant tree species in the canopy layer (upper tree layer). In addition to the canopy tree species, the data collection recorded different types of canopy gaps such as grass and shrub-covers, bare soil, and recently burned patches (i.e. burned scars) along each survey transect. The tree species that were sampled are listed in Table 1.

2.3. Spectral measurement and wet chemical analysis

In total, 82 and 78 leaf samples were collected from the Dukuduku Tropical Forest and Savanna landscape, respectively. These samples were then subjected to spectral measurement and wet chemical analysis. The spectral reflectance and carotenoids contents of leaves were measured within three days of sampling using an ASD FieldSpec3(R) spectrometer equipped with a leaf contact probe. A spectralon panel was used as a reference standard for the spectral measurements. The spectrometer provides a spectral range of 350 to 2500 nm with a 1 nm resolution, covering the visible, near infrared, and shortwave infrared regions of the electromagnetic spectrum. For each leaf sample, two reflectance measurements were taken and averaged to account for spatial

Table 1. Leaf datasets.

Ecosystem type	Species name	No. of points Spectral sample per species	Average carotenoids content and standard deviation
savanna trees	Leaves of different ages for 8 species (*Combretum hereroense*, *Combretum molle*, *Combretum collinum*, *Euclea natalensis*, *Terminalia sericea*, *Sclerocarya birrea*, *Pterocarpus rotundifolius* and *Lannea discolour*)	78	Greater Kruger National Park area, Mpumalanga Province, South Africa.
tropical trees	Species of broadleaved trees, coastal forest (*Albizia adianthifolia, Strchynos spp, Acacia, Ziziphus mucronate, Trichilia dregeana, Strychnos spp, Lannea spp, Ficus spp, Celtis Africana, Balanites maughamii*, and *Bridelia micrantha)*	82	St Lucia, South Africa

heterogeneity within the leaf structure. The contact probe had a 25 mm diameter and a 10 mm instantaneous field of view, with its own halogen lamp light source used for the measurements (Main et al. 2011). It is important to note that the full spectral range was utilized in two ways in this study. Firstly, due to the unique absorption properties of carotenoids in the visible portion of the spectrum, this study focused on analysing the visible region of the spectrum (400–745 nm). Second, various machine learning algorithms, including random forest, XGBoost, partial least-squares regression (PLSR), and artificial neural networks (ANN), were used to perform feature selection across the entire spectral range (i.e. 2100 bands). The wavelengths identified as most important by the best-performing algorithm were then further incorporated into the PROSPECT-5 RTM inversion.

After the spectral measurements, the leaf samples were used for wet chemical analysis to determine carotenoid concentration. Here, a leaf borer with a diameter of 18 mm was used to collect a sample from the same area of the leaf that had been measured. The collected leaf samples were carefully maintained in a cool and dry state before being transported for carotenoid content analysis within 24 hours. The wet lab extraction technique was employed to determine the carotenoid concentration per unit area of the leaf (Lichtenthaler and Wellburn 1983). After recording the leaves fresh weight, the leaf pigments were extracted using 100% acetone as per the equations below.

$$C_a = 11.24A_{661.2} - 2.04A_{644.8} \tag{1}$$

$$C_b = 20.13A_{644.8} - 4.19A_{661.2} \tag{2}$$

$$C_{x+c=}1000A_{470} - 2.27C_a - 81.4C_b/227 \tag{3}$$

2.4. Statistical analysis

The ASD FieldSpec generate high-dimensional data due to its ability to capture many spectral bands (i.e. 2100 bands). Due to the high correlation between independent and dependent variables, this poses a challenge for many regression analyses, such as multicollinearity. Dealing with such high-dimensional hyperspectral data requires a robust algorithm to handle its unique characteristics. Thus, Random Forest regression (RFR), XGBoost, Partial Least Square Regression (PLSR) and Artificial Neural Network (ANN) were tested to estimate carotenoid content.

2.4.1. Random forest regression

In this study, regression analysis was conducted using the random forest algorithm to identify wavelength informative for estimating carotenoids. Random forest is a popular machine-learning technique for feature selection and prediction. In the context of carotenoid analysis, random forests can help identify the specific wavelengths most informative for predicting carotenoid concentrations. The random forest algorithm works by constructing an ensemble of decision trees, where each tree is trained on a random subset of the available wavelengths (Ließ, Schmidt, and Glaser 2016; Mngadi, Odindi, and Mutanga 2021). During the training process, the algorithm assesses the importance of

each wavelength by measuring how much it contributes to the model's overall accuracy or predictive power. The importance of each wavelength is typically measured using metrics such as Gini importance or Mean Decrease Impurity (Breiman 2001).

2.4.2. XGBoost

XGBoost regression is an advanced gradient-boosting algorithm widely used for regression tasks. XGBoost leverages the collective strength of numerous individual weak decision trees to generate a robust and accurate predictive model (Chen and Guestrin 2016). XGBoost regression is important for feature selection because it can automatically rank the importance of features based on their contribution to the overall predictive performance. This algorithm operates through an iterative process of constructing decision trees, with each subsequent tree aimed at rectifying the errors made by the previous ones (Hengl et al. 2017). During this iterative process, XGBoost assigns weights to features based on their ability to minimize the objective function, which measures the model's error. By examining the accumulated weights of features across multiple trees, XGBoost identifies the most important features that impact prediction accuracy. Consequently, it enables efficient feature selection and facilitates more accurate regression modelling.

2.4.3. Partial Least Square Regression (PLSR)

PLSR is a well-established technique that combines the dependent and independent variables to select the optimal variables (Song et al. 2011). In this case, PLS makes it possible to find spectral bands closely correlated to carotenoids and flavonoids. PLS regression is a superior method because the analysis is not done band-by-band, but a continuous spectrum is treated as a single measurement. Also, the major advantage of PLS compared to Principal Component Analysis (PCA) is that the dependent variable is considered when performing dimensionality reduction and eliminating overlapping wavelengths (Lehmann et al. 2015). Both independent variables (X) and dependent variables (Y) are reduced to principal components (PCs). Once the PLS has been performed, several components are formed, each explaining a portion of the variance.

2.4.4. Artificial Neural Network (ANN)

Artificial neural networks are computational models that draw inspiration from the structure and functioning of the human brain. These networks are composed of interconnected nodes, often called 'neurons', organized into layers. Each neuron carries out weighted computations and transfers the results to the subsequent layer. ANNs are important for feature selection because they can automatically learn and extract relevant patterns and relationships from complex data, enabling the identification of the most informative features (Xu, Wang, and Shi 2020). By leveraging their ability to capture nonlinear relationships and adapt to different data types, ANNs facilitate efficient and effective feature selection, improving model performance, interpretability, and computational efficiency.

The best-performing algorithm wavelengths were further incorporated in the PROSPECT-5 inversion. The model's accuracy was assessed by comparing predicted values to the measured in the test set. Validation involved calculating the root mean squared error (RMSE) and coefficient of determination (R-squared). The entire dataset ($n = 160$) was divided, allocating 70% as the training set and 30% as the testing dataset.

2.5. Simulated datasets using the PROSPECT-5 model

Utilizing the PROSPECT-5 model, 100,000 randomly simulated spectra were generated using six input variables: leaf structure parameter (N), LChl, LCar, brown pigments (Cbrown), equivalent water thickness (EWT), and Leaf mass per area (LMA). Typically, pigments absorb light between 400 and 700 nm in the visible spectrum, water has a strong absorption in the 1000–2500 nm near-infrared, and dry matter is present throughout the entire wave range. Therefore, the EWT value was set constant as this study sought to measure carotenoid content in the visible range, where EWT did not influence the outcome. Hence, a look-up database was constructed using the 100,000 spectra and matching parameter value sets. For the experiment, six optical domains that include, visible (400–700 nm), blue (450–510 nm), green (510–580 nm), yellow (585–625 nm), red (630–690 nm), red-edge (705–745 nm), and bands were selected using random forest (Figure 2a). Table 2 displays the input parameter values employed in the PROSPECT-5 (Feret et al. 2008) simulations. Therefore, in this study, we used Artificial Neural Network (ANN) inversion methods on the selected wavelengths to determine a robust, accurate, and dependable approach for the inversion of PROSPECT-5 from ASD data. The neural network with three layers (input, hidden, and output layer) was utilized in Jupyter Notebook using the sklearn package in the Python 2.7 version.

The precision of the retrievals was demonstrated through the comparison of the *in-situ* data and the prediction using coefficient of determination (R^2) (Equation 4) and root mean square error (RMSE) (Equation 5).

$$R^2 = 1 - \frac{\sum^{(y_i - y_i')}2}{\sum^{(y_i - \bar{y_i})}2} \tag{4}$$

$$RMSE = \sqrt{\frac{\sum^{(y_i - y_i')}2}{n}} \tag{5}$$

3. Results

3.1. Algorithms prediction performance

Among the four tested models, the random forest was the most robust in estimating carotenoids, indicating an R^2 of 0.82 and RMSE of 1.436. This was closely followed by XGBoost ($R^2 = 0.80$ and RMSE = 1.501) and ANN ($R^2 = 0.80$ and RMSE = 1.641). Lastly, PLSR was the least robust, with an R^2 of 0.75 and RMSE of 1.897. Figure 1 illustrates the correlation between the observed and the estimated carotenoids for the constructed models.

Table 2. Input parameters of PROSPECT model utilized for generating the LUT and their ranges.

			Range	
Parameter	Symbol	Unit	Min	Max
Leaf structure parameter	(N)	NA	1.2	1.2
Leaf chlorophyll content	LChl	µg/cm2	20	80
Leaf carotenoid content	LCar	µg/cm2	2	10
Equivalent water thickness	EWT	g/cm2	0.021	–
Leaf mass per area	LMA	g/cm2	0.001	0.016
Brown	Cbrown	NA	0	0

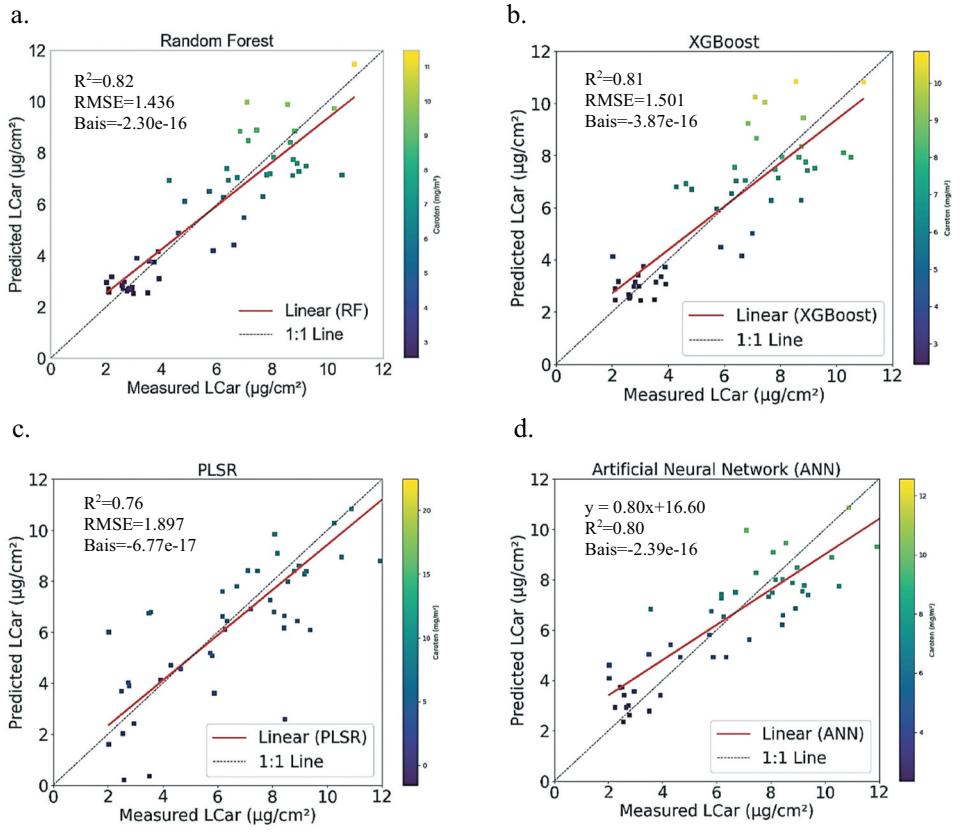

Figure 1. Predicted against observed carotenoids using various algorithms.

3.2. Variable importance selection using the random forest model

For demonstration purposes, Figure 2 illustrates the most influential wavelength by various algorithms. However, since the random forest model was the most robust, we focused on the important wavelengths of the model to estimate carotenoids (Figure 2a). The most influential wavelength ranges include the yellow (581, 587, 589, 595, 599, 600, 605, 606, 609, 610, 617), green (557, 558, 570), and red-edge (703, 708) region of the spectrum. The selected bands by random forest were further tested in the Prospect 5 model to constrain the inversion.

3.3. Results of PROSPECT-5 model for savanna, tropical forest, and combined datasets

Figure 3 illustrates the root mean square error (RMSE) between measured and simulated reflectance for the savanna, tropical forest, and combined datasets, respectively. The results indicate that the tropical forest dataset produced the lowest RMSE among the three, with an average RMSE of approximately 1.55 and a maximum value below 2.58. This was followed by the combined dataset, which had an average RMSE of 1.86 and a maximum value of 2.65. The savanna dataset exhibited the highest RMSE, with an average of 1.98 and a maximum of 2.85.

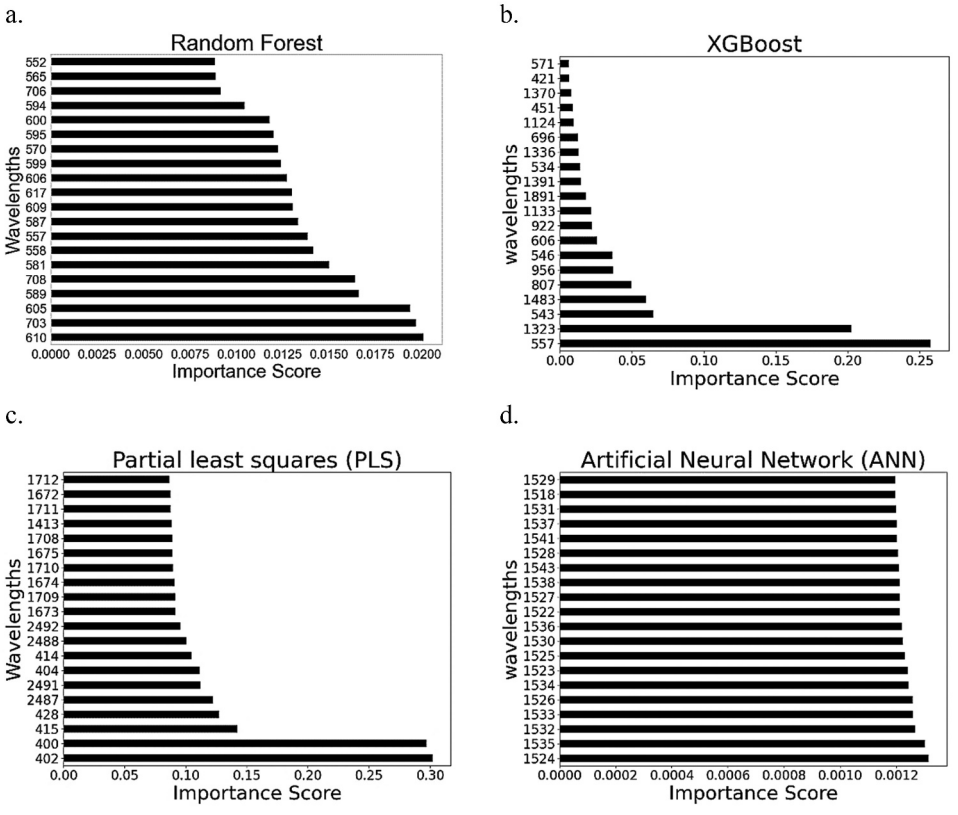

Figure 2. Results of various algorithms in introducing the most important wavelengths.

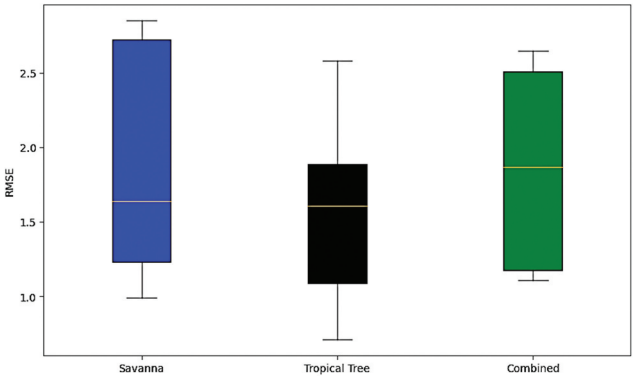

Figure 3. Root mean square error (RMSE) of the three datasets (savanna, tropical forest, and combined) between measured and simulated spectra using PROSPECT-5.

To further investigate the factors influencing the performance of the different models, the spectral regions with the lowest RMSE among the three datasets were identified and the simulated and measured spectra compared (Figure 4). The findings reveal that the green region had the lowest RMSE of 0.71 across all models

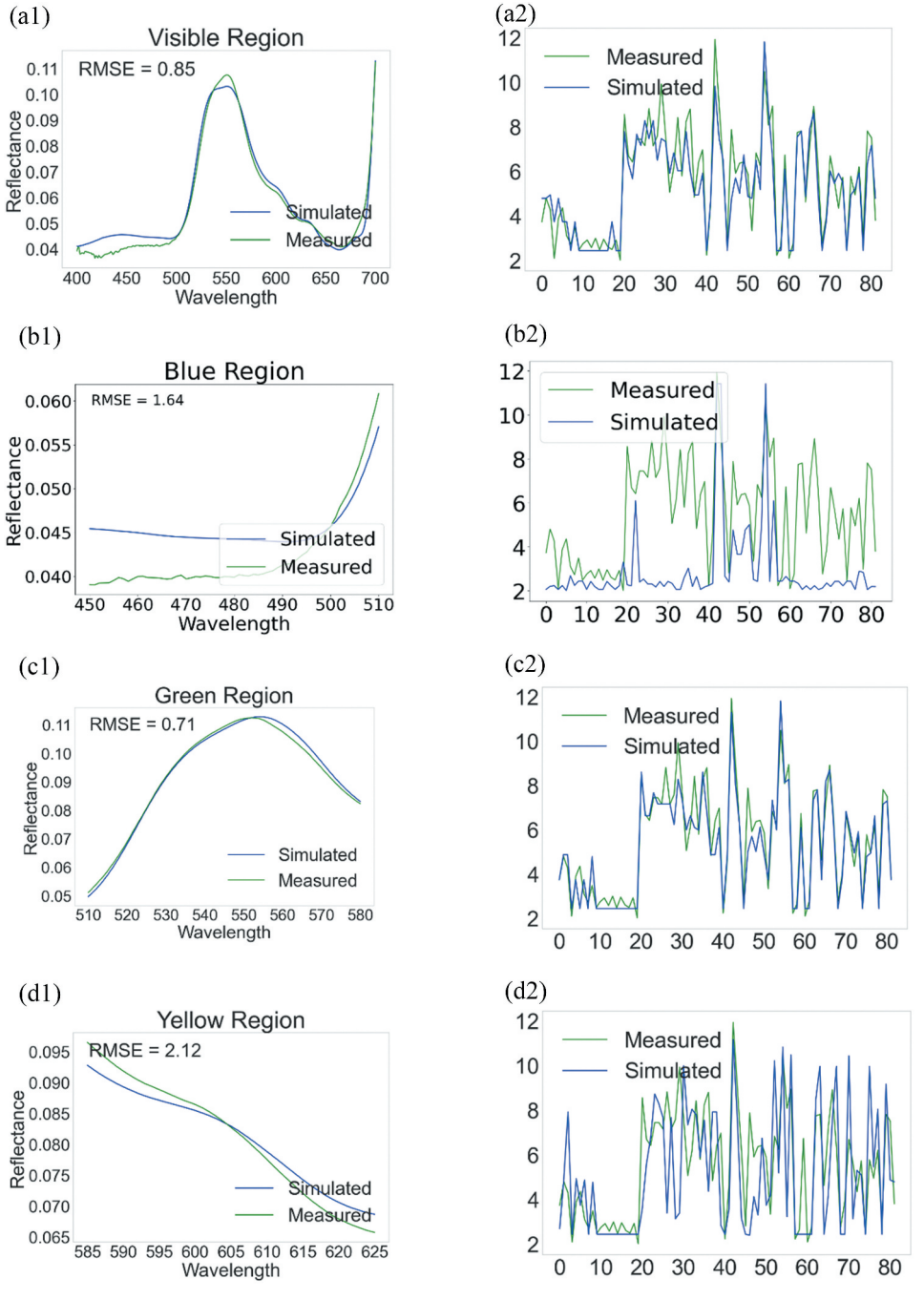
(a1)
Visible Region
RMSE = 0.85
Reflectance
Wavelength
Simulated
Measured
(a2)
Measured
Simulated
(b1)
Blue Region
RMSE = 1.64
Reflectance
Wavelength
Simulated
Measured
(b2)
Measured
Simulated
(c1)
Green Region
RMSE = 0.71
Reflectance
Wavelength
Simulated
Measured
(c2)
Measured
Simulated
(d1)
Yellow Region
RMSE = 2.12
Reflectance
Wavelength
Simulated
Measured
(d2)
Measured
Simulated

Figure 4a.

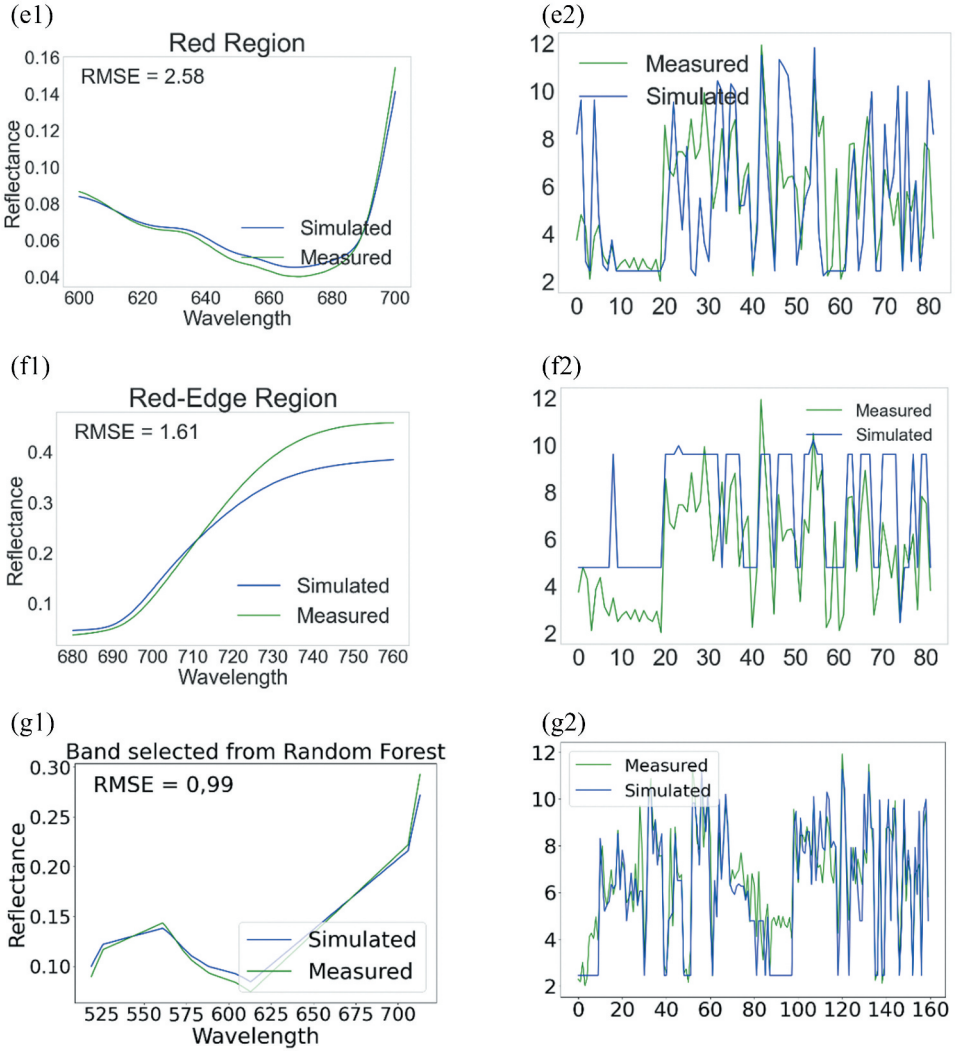

Figure 4b. Spectral matching between simulated PROSPECT-5 and measured spectral, and differences between simulated and measured spectra.

(Figure 4c1–2), followed by the visible region with an RMSE of 0.85 (Figure 4a1–2). The bands selected by the random forest algorithm had an RMSE of 0.99 (Figure 4g1–2), respectively. Among the selected bands with the lowest RMSE, the majority were from the tropical forest dataset, including the visible, green, yellow, red, and red-edge regions. The blue region and the bands selected by the random forest algorithm were derived from the savanna dataset. The analysis of individual performance among the three datasets, which includes Savanna, tropical tree and combined, respectively, is shown in the next section.

3.3.1. Savanna

The analysis of the savanna landscape using various spectral regions (visible, blue, green, yellow, red, and red-edge), and the bands selected from random forest showed varying performance levels. The key findings are as follows: the bands selected by the random forest algorithm demonstrated the highest overall performance, with an R^2 of 0.84 and a Root Mean Squared Error (RMSE) of 0.99 (Figure 5g). This approach outperformed the other spectral regions, suggesting that the band selected from random forest is the most effective in estimating carotenoids. This was followed by the green region which also exhibited strong performance, with an overall accuracy of $R^2 = 0.70$ and an RMSE of 1.19 (Figure 5c). The visible region produced a respectable overall accuracy of $R^2 = 0.68$ and an RMSE of 1.27 (Figure 5a).

On the contrary, the blue region showed the lowest performance among all the models, with an overall accuracy of $R^2 = 0.17$ and an RMSE of 1.64 (Figure 5b), closely followed by the red region with an overall accuracy of $R^2 = 0.38$ (Figure 5e). These results highlight the importance of carefully selecting the appropriate spectral regions for analysing savanna landscapes, as different regions can perform differently. The band selected from random forest demonstrated the highest overall accuracy and the lowest RMSE, making it a potentially valuable approach to estimating carotenoids.

3.3.2. Tropical forest

The analysis of the Dukuduku indigenous tropical forest using various spectral regions and bands selected from random forest algorithm shows that the green region had the highest overall performance, with an R^2 of 0.90 and an RMSE of 0.71 (Figure 6c). The visible region closely followed the green region, with an overall accuracy of $R^2 = 0.84$ and an RMSE of 0.85 (Figure 6a). The band selected from random forest had an overall accuracy of $R^2 = 0.78$ and an RMSE of 1.33, which places it as the third-best performing approach in this context (Figure 6g).

On the other hand, the blue region showed the lowest performance, with an overall accuracy of $R^2 = 0.20$ and an RMSE of 1.66 (Figure 6b), and was closely followed by the red region with an overall accuracy of $R^2 = 0.30$ and an RMSE of 2.58 (Figure 6e). These results highlight the importance of carefully selecting spectral regions for analysing tropical forest landscapes, as different regions can perform differently. The green region demonstrated the highest overall accuracy and the lowest RMSE, making it a valuable approach for this type of analysis.

3.3.3. Combined

The study further evaluated the performance of both a savanna landscape and the Dukuduku tropical forest. By combining the results from these two distinct ecosystems, the analysis provides a comprehensive assessment of the efficacy of the different spectral domains. The band selected from the random forest algorithm demonstrated the highest overall performance, with an R^2 of 0.80 and a Root Mean Squared Error (RMSE) of 1.20 (Figure 7g). The green region also exhibited strong performance, with an overall accuracy of $R^2 = 0.75$ and an RMSE of 1.11 (Figure 7c). The visible region produced a respectable overall accuracy of $R^2 = 0.72$ and an RMSE of 1.15 (Figure 7a).

Lastly, the blue region showed the lowest performance, with an overall accuracy of $R^2 = 0.20$ and an RMSE of 1.87 (Figure 7b) and was closely followed by the red-edge

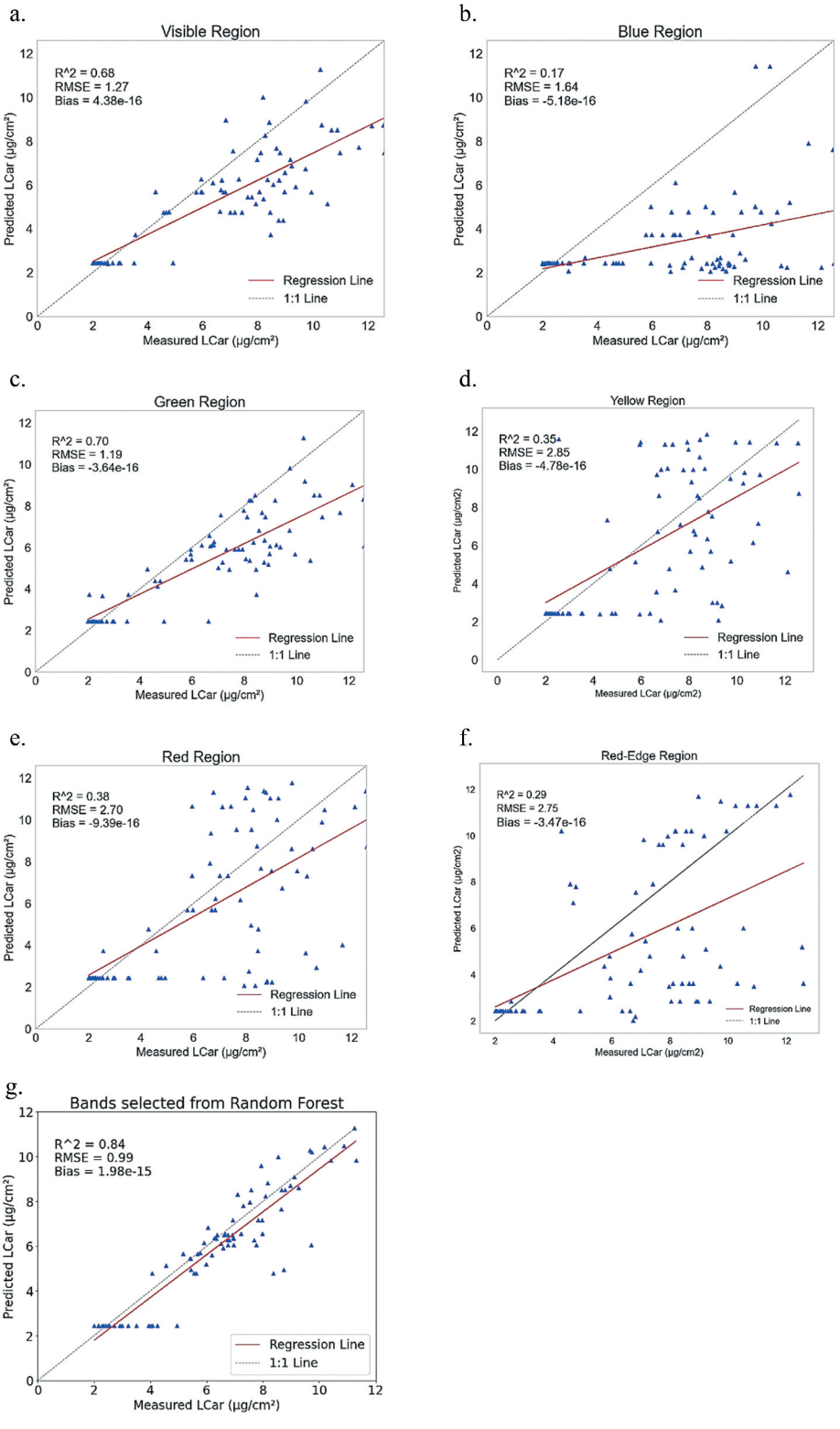

Figure 5. Relationships between observed and predicted carotenoids (µg/cm^2) for various spectral regions and bands selected from random forests algorithm in the savanna landscape, derived from PROSPECT-5 inversion.

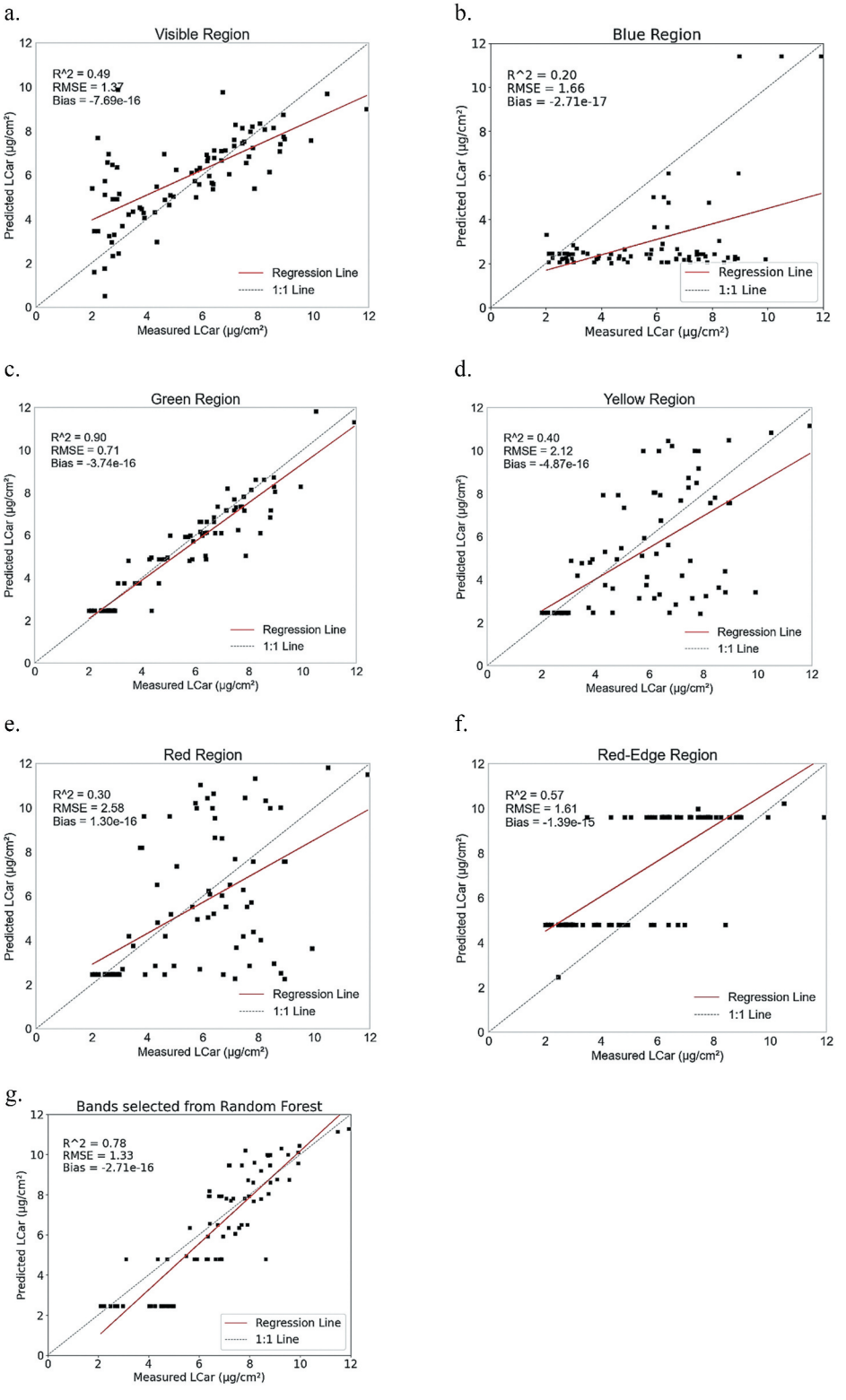

Figure 6. Relationships between observed and predicted carotenoids in the dukuduku tropical forest using different spectral regions and bands selected from random forests.

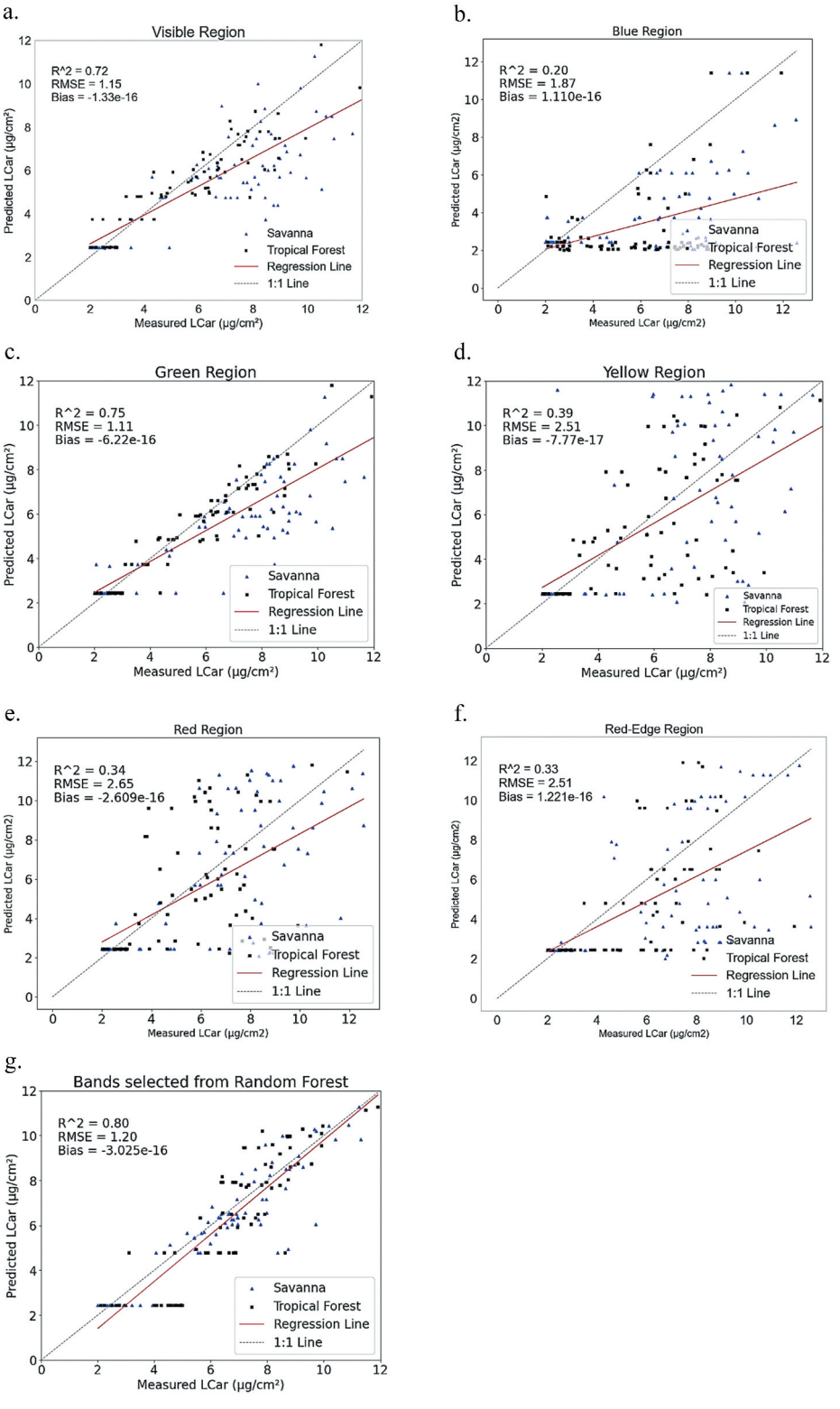

Figure 7. Relationships between observed and predicted carotenoids (µg/cm^2) for various spectral regions and bands selected from random forests algorithm for both savanna landscape and tropical forest dataset, derived from PROSPECT-5 inversion.

region which had an overall accuracy of $R^2 = 0.33$ and an RMSE of 2.51 (figure f). These results highlight the importance of carefully selecting the appropriate spectral regions for analysing both savanna landscapes and tropical forests, as different regions can perform differently. The band selected from random forest algorithm demonstrated the highest overall accuracy and the lowest RMSE, making it a potentially valuable approach for this type of analysis.

Overall, these findings highlight the importance of carefully evaluating the performance of different spectral regions and models when estimating carotenoid content in both savanna and tropical forest environments. The band selected from random forest, the green and visible region of spectra emerged as the most promising approaches for these applications.

4. Discussion

Traditionally, PROSPECT model uses the entire domain reflectance (400–2500 nm) to estimate leaf contents. However, using waveband characteristics can significantly reduce the need for complete domain reflectance and improve the inversion performance. In this study, we proposed a wavelength selection approach utilizing PROSPECT5-ANN models to estimate carotenoids.

The analysis of three distinct datasets, encompassing savanna, tropical forest, and a combination of the two (both savanna and tropical forest) revealed that the green region demonstrated the strongest overall performance. This is evidenced by an R^2 of 0.90 and a RMSE of 0.71, which were observed in the tropical forest dataset. This high correlation between the green reflectance and carotenoid content is consistent with findings from numerous studies that have highlighted the green region as a sensitive indicator of carotenoid dynamics across a range of plant functional types and biomes (Gamon, Serrano, and Surfus 1997; Gitelson et al. 2002; Sims and Gamon 2002; Stylinski, Gamon, and Oechel 2002). Studies by Gitelson et al. (2002), Blackburn (2007), and Gamon, Serrano, and Surfus (1997) for instance demonstrated similar strong correlation between carotenoid levels and reflectance in the green spectral region. These and other works have helped establish the green region as a sensitive indicator of carotenoid status in plants, with applications in remote sensing, plant stress monitoring, and other vegetation applications.

The strong correlation between carotenoid levels and green reflectance observed in the tropical forest dataset can be attributed to the relatively stable leaf pigment composition and optical properties in the evergreen ecosystem. Hence, the tropical forest showed the lowest RMSE among all three tested data to estimate carotenoids (Figure 3). Furthermore, the high and consistent chlorophyll content in the Dukuduku forest may reduce the interference of other pigments (e.g. anthocyanins) on the green reflectance signal, allowing the green region to directly capture the variations in carotenoid content. In the savanna ecosystem, the green region also exhibited the second-best performance, with an R^2 of 0.70 and an RMSE of 1.19. A similar trend was observed in the combined dataset, where the green region again demonstrated the second-best performance, with an R^2 of 0.72 and an RMSE of 1.11. These findings further corroborate the importance of the green region as a sensitive indicator of carotenoid dynamics across different biomes.

The bands selected by the random forest model produced the second-highest accuracy and this was observed in the savanna ecosystem, with an R^2 of 0.84 and RMSE of 0.99. For the combined dataset, the band selected by random forest model produced the highest accuracy, while for the Dukuduku tropical forest, it produced the third-highest accuracy. The high performance of the bands chosen by the random forest model may be attributed to the ability of the random forest algorithm to gain a holistic understanding of the interconnections between vegetation attributes throughout the entire spectrum, allowing for the identification of wavelengths that contain more conclusive information for estimating carotenoid content (Breiman 2001; Ghosh and Behera 2018). The wavelengths selected by the random forest model included the green, yellow, and red-edge regions of the visible spectrum, which aligns with this study's findings, suggesting the visible region's importance for carotenoid estimation.

The visible region produced the second-highest accuracy for the Dukuduku tropical forest dataset ($R = 0.84$ and RMSE = 0.85) and the combined dataset, ($R = 0.72$ and RMSE = 1.15) and the third-highest performance for the savanna ecosystem ($R = 0.68$ and RMSE = 1.27). This suggests that the visible region may be a suitable alternative for estimating carotenoids, particularly in situations where more advanced modelling techniques like random forests algorithm are not feasible or accessible. The visible region has been used extensively in literature for retrieving carotenoid content from remotely sensed data, as carotenoids influence the reflectance characteristics in this region of the spectrum (Blackburn 2007; Gamon, Serrano, and Surfus 1997; Gitelson and Merzlyak 2004; Sonobe and Wang 2018). The visible region demonstrated strong performance in estimating carotenoid content across the Dukuduku tropical forest and combined (Dukuduku and savanna) datasets, achieving the second-highest accuracy. This suggests that the visible region effectively captures the distinct optical signatures created by varying carotenoid concentrations in the leaves, likely due to the high and consistent chlorophyll content in the evergreen Dukuduku ecosystem reducing interference from other pigments.

The visible region's sensitivity to carotenoid content remained robust even when integrating data from contrasting vegetation types, indicating its broad applicability across diverse biomes. In the savanna ecosystem, the visible region exhibited the third-highest performance. While not as strong as in the other datasets, the visible region's performance was still considerable, highlighting its potential for use in savanna environments. It is important to note that while the savanna dataset exhibited considerable accuracy in estimating carotenoid content, it was not as strong as the performance of the tropical forest dataset and combined, respectively. The savanna dataset produced the highest RMSE among the three datasets (Figure 3). This can be attributed to the landscape's complex structure.

Savannas are characterized by an open herbaceous layer of grass species and an intermittent layer of trees. This heterogeneous arrangement, with varying levels of vegetation cover and composition, presents a more challenging environment for the PROSPECT-5 model to accurately capture the spectral signatures related to carotenoid content (Madonsela 2018). The variable mix of grasses and scattered trees creates a complex radiative transfer environment that may not be as well represented by the PROSPECT-5 model's assumptions and parameterizations. In contrast, the Dukuduku tropical forest is composed of a continuous, evergreen canopy layer, which likely provides

a more consistent and predictable spectral response for the PROSPECT-5 model to accurately simulate carotenoid content. The homogenous and dense vegetation structure of the tropical forest appears to be better suited for the PROSPECT-5 model's performance compared to the more heterogeneous savanna landscape.

The findings reveal a consistent pattern of poor performance by the blue and red spectral regions across the different datasets and models examined. For the combined dataset, the blue region exhibited the lowest overall accuracy ($R^2 = 0.20$) and the highest RMSE (1.87), closely followed by the red-edge region ($R^2 = 0.33$, RMSE = 2.51). A similar trend was observed in the savanna ecosystem, where the blue region again showed the lowest performance ($R^2 = 0.20$, RMSE = 1.66), followed by the red region ($R^2 = 0.30$, RMSE = 2.58). The poor performance of the blue and red spectral regions could be attributed to the vegetation characteristics of the studied ecosystems. In Dukuduku tropical forest, the evergreen nature of the vegetation likely results in high chlorophyll content, which would lead to strong absorption of blue and red wavelengths. Similarly, in the savanna ecosystem, if the data was collected during the summer season, the vegetation may have been at its peak greenness, with a high chlorophyll content. Under such conditions, the blue and red wavelengths would be heavily absorbed, limiting their ability to capture the nuances of carotenoid variations. In contrast, the visible region, which encompasses the green wavelengths, may have been better suited to detect the distinct optical signatures created by varying carotenoid concentrations, even in the presence of high chlorophyll levels. The influence of seasonal and ecosystem-specific characteristics on the performance of different spectral regions highlights the importance of understanding the underlying biophysical and biochemical processes governing vegetation reflectance, which can help optimize the selection of spectral bands and the development of more robust remote sensing-based approaches for monitoring vegetation properties such as carotenoid content across diverse biomes.

Based on this study's findings, we recommend using machine learning algorithms such as random forest and visible sub-spectral domain to select characteristic wavelengths, instead of using the full spectral domain when performing the inversion for carotenoids. The PROSPECT model can accurately estimate carotenoid concentrations by focusing on the band selection method, which aligns with the carotenoids absorption bands. The visible range is particularly relevant for carotenoids, as they exhibit strong absorption peaks within this region. Therefore, restricting the analysis to the bands selected by random forest or visible spectrum enhances the model's sensitivity to carotenoid content and improves the inversion results.

The PROSPECT-5 model has demonstrated promising results in simulating carotenoid content at the leaf level. However, to further investigate this approach, it is necessary to evaluate its performance at the canopy scale. At the canopy level, additional parameters need to be considered, such as leaf area, soil properties, leaf angle, and various angular effects (observer zenith, relative azimuth, solar zenith, and hotspot). To account for these canopy-level factors, the PROSPECT model can be combined with a canopy model, such as SAIL, to create the PROSAIL framework. This integrated approach can provide a more comprehensive understanding of the interactions between leaf-level properties and canopy-level factors, and their collective influence on the remote sensing signal.

Future studies should test the proposed wavelength selection approach within the PROSAIL framework to evaluate its performance in estimating carotenoid content at the canopy scale, where the impacts of noise and other confounding factors can be more realistically simulated. Additionally, existing carotenoid-related indices and their first-order derivatives should be tested on the current dataset across different ecosystems. This cross-validation and comparative analysis can help identify the most robust and versatile methods for carotenoid estimation in diverse biomes.

5. Conclusion

In this study, the wavelength selection approach was applied to estimate carotenoids using the PROSPECT-5. The following conclusions can be drawn:

- Among the tested algorithms for feature selection, random forest surpassed all other models with an R^2 of 0.82 and RMSE of 14.36.
- The analysis of three distinct datasets, encompassing savanna, tropical forest, and a combination of the two (both savanna and tropical forest) revealed that the green region, the band selected by the random forest, and visible region outperformed all other models.

Overall, this study is the first to use the band selection method with the PROSPECT model to simulate carotenoids. The results are valuable in determining the methods or bands valuable in estimating carotenoids.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The work was supported by the National Research Foundation (NRF) [NEO960912 SIBIYA].

ORCID

Bongokuhle Sibiya http://orcid.org/0000-0002-0058-5390
John Odindi http://orcid.org/0000-0002-4934-1346

References

Asner, G. P., M. Palace, M. Keller, R. Pereira Jr, J. N. Silva, and J. C. Zweede. 2002. "Estimating Canopy Structure in an Amazon Forest from Laser Range Finder and IKONOS Satellite Observations 1." *Biotropica* 34 (4): 483–492. https://doi.org/10.1111/j.1744-7429.2002.tb00568.x.

Blackburn, G. A. 1998. "Spectral Indices for Estimating Photosynthetic Pigment Concentrations: A Test Using Senescent Tree Leaves." *International Journal of Remote Sensing* 19 (4): 657–675. https://doi.org/10.1080/014311698215919.

Blackburn, G. A. 2007. "Hyperspectral Remote Sensing of Plant Pigments." *Journal of Experimental Botany* 58 (4): 855–867.

Breiman, L. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32. https://doi.org/10.1023/A:1010933404324.

Chappelle, E. W., M. S. Kim, and J. E. McMurtrey III. 1992. "Ratio Analysis of Reflectance Spectra (RARS): An Algorithm for the Remote Estimation of the Concentrations of Chlorophyll A, Chlorophyll B, and Carotenoids in Soybean Leaves." *Remote Sensing of Environment* 39 (3): 239–247. https://doi.org/10.1016/0034-4257(92)90089-3.

Chen, T., and C. Guestrin. 2016, August. "Xgboost: A Scalable Tree Boosting System." In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785–794.

Cho, M. A., O. Malahlela, and A. Ramoelo. 2015. "Assessing the Utility WorldView-2 Imagery for Tree Species Mapping in South African Subtropical Humid Forest and the Conservation Implications: Dukuduku Forest Patch as Case Study." *International Journal of Applied Earth Observation and Geoinformation* 38:349–357. https://doi.org/10.1016/j.jag.2015.01.015.

Cho, M. A., R. Mathieu, G. P. Asner, L. Naidoo, J. A. N. Van Aardt, A. Ramoelo, P. Debba, et al. 2012. "Mapping Tree Species Composition in South African Savannas Using an Integrated Airborne Spectral and LiDAR System." *Remote Sensing of Environment* 125:214–226. https://doi.org/10.1016/j.rse.2012.07.010.

Cho, M. A., A. Ramoelo, and A. Skidmore. 2014, July. "Exploring Various Spectral Regions for Estimating Chlorophyll from ASD Leaf Reflectance Using Prospect Radiative Transfer Model." In *2014 IEEE Geoscience and Remote Sensing Symposium*, 4754–4757. IEEE.

Datt, B. 1998. "Remote Sensing of Chlorophyll A, Chlorophyll B, Chlorophyll A+ B, and Total Carotenoid Content in Eucalyptus Leaves." *Remote Sensing of Environment* 66 (2): 111–121. https://doi.org/10.1016/S0034-4257(98)00046-7.

Feret, J. B., C. François, G. P. Asner, A. A. Gitelson, R. E. Martin, L. P. Bidel, S. L. Ustin, G. Le Maire, and S. Jacquemoud. 2008. "PROSPECT-4 and 5: Advances in the Leaf Optical Properties Model Separating Photosynthetic Pigments." *Remote Sensing of Environment* 112 (6): 3030–3043. https://doi.org/10.1016/j.rse.2008.02.012.

Féret, J. B., A. A. Gitelson, S. D. Noble, and S. Jacquemoud. 2017. "PROSPECT-D: Towards Modeling Leaf Optical Properties Through a Complete Lifecycle." *Remote Sensing of Environment* 193:204–215. https://doi.org/10.1016/j.rse.2017.03.004.

Gamon, J., L. Serrano, and J. S. Surfus. 1997. "The Photochemical Reflectance Index: An Optical Indicator of Photosynthetic Radiation Use Efficiency Across Species, Functional Types, and Nutrient Levels." *Oecologia* 112 (4): 492–501. https://doi.org/10.1007/s004420050337.

Ghosh, S. M., and M. D. Behera. 2018. "Aboveground Biomass Estimation Using Multi-Sensor Data Synergy and Machine Learning Algorithms in a Dense Tropical Forest." *Applied Geography* 96:29–40. https://doi.org/10.1016/j.apgeog.2018.05.011.

Gitelson, A. A., G. P. Keydan, and M. N. Merzlyak. 2006. "Three-Band Model for Noninvasive Estimation of Chlorophyll, Carotenoids, and Anthocyanin Contents in Higher Plant Leaves." *Geophysical Research Letters* 33 (11). https://doi.org/10.1029/2006GL026457.

Gitelson, A. A., and M. N. Merzlyak. 2004. "Non-Destructive Assessment of Chlorophyll Carotenoid and Anthocyanin Content in Higher Plant Leaves: Principles and Algorithms."

Gitelson, A. A., Y. Zur, O. B. Chivkunova, and M. N. Merzlyak. 2002. "Assessing Carotenoid Content in Plant Leaves with Reflectance spectroscopy." *Photochemistry and Photobiology* 75 (3): 272–281. https://doi.org/10.1562/0031-8655(2002)075<0272:ACCIPL>2.0.CO;2.

He, C., J. Sun, Y. Chen, L. Wang, S. Shi, F. Qiu, S. Wang, J. Yang, and T. Tagesson. 2023. "PROSPECT-GPR : Exploring Spectral Associations Among Vegetation Traits in Wavelength Selection for Leaf Mass per Area and Water Contents." *Science of Remote Sensing* 8:100100. https://doi.org/10.1016/j.srs.2023.100100.

Hengl, T., J. G. Leenaars, K. D. Shepherd, M. G. Walsh, G. B. Heuvelink, T. Mamo, H. Tilahun, et al. 2017. "Soil Nutrient Maps of Sub-Saharan Africa: Assessment of Soil Nutrient Content at 250 M Spatial Resolution Using Machine Learning." *Nutrient Cycling in Agroecosystems* 109 (1): 77–102. https://doi.org/10.1007/s10705-017-9870-x.

Hernández-Clemente, R., R. M. Navarro-Cerrillo, and P. J. Zarco-Tejada. 2012. "Carotenoid Content Estimation in a Heterogeneous Conifer Forest Using Narrow-Band Indices and PROSPECT+ DART

Simulations." *Remote Sensing of Environment* 127:298–315. https://doi.org/10.1016/j.rse.2012.09.014.

Huang, W., X. Zhou, W. Kong, and H. Ye. 2018. "Monitoring Crop Carotenoids Concentration by Remote Sensing." In *Progress in Carotenoid Research*, London, UK: IntechOpen.

Lawes, M. J., D. M. Macfarlane, and H. A. Eeley. 2004. "Forest Landscape Pattern in the KwaZulu–Natal Midlands, South Africa: 50 Years of Change or Stasis?" *Austral Ecology* 29 (6): 613–623.

Lehmann, J. R. K., A. Große-Stoltenberg, M. Römer, and J. Oldeland. 2015. "Field Spectroscopy in the VNIR-SWIR Region to Discriminate Between Mediterranean Native Plants and Exotic-Invasive Shrubs Based on Leaf Tannin Content." *Remote Sensing* 7 (2): 1225–1241. https://doi.org/10.3390/rs70201225.

Li, D., T. Cheng, M. Jia, K. Zhou, N. Lu, X. Yao, Y. Tian, Y. Zhu, and W. Cao. 2018. "PROCWT: Coupling PROSPECT with Continuous Wavelet Transform to Improve the Retrieval of Foliar Chemistry from Leaf Bidirectional Reflectance Spectra." *Remote Sensing of Environment* 206:1–14. https://doi.org/10.1016/j.rse.2017.12.013.

Lichtenthaler, H. K., and A. R. Wellburn. 1983. "Determinations of Total Carotenoids and Chlorophylls a and B of Leaf Extracts in Different Solvents." *Biochemical Society Transactions* 11 (5): 591–592. https://doi.org/10.1042/bst0110591.

Ließ, M., J. Schmidt, and B. Glaser. 2016. "Improving the Spatial Prediction of Soil Organic Carbon Stocks in a Complex Tropical Mountain Landscape by Methodological Specifications in Machine Learning Approaches." *PLOS ONE* 11 (4): e0153673.

Lu, B., C. Proctor, and Y. He. 2021. "Investigating Different Versions of PROSPECT and PROSAIL for Estimating Spectral and Biophysical Properties of Photosynthetic and Non-Photosynthetic Vegetation in Mixed Grasslands." *GIScience & Remote Sensing* 58 (3): 354–371. https://doi.org/10.1080/15481603.2021.1877435.

Madonsela, S. 2018. *Exploring the Relationship Between Spectral Reflectance and Tree Species Diversity in the Savannah Woodlands* (Doctoral dissertation).

Madonsela, S., M. A. Cho, A. Ramoelo, and O. Mutanga. 2017. "Remote Sensing of Species Diversity Using Landsat 8 Spectral Variables." *Isprs Journal of Photogrammetry & Remote Sensing* 133:116–127. https://doi.org/10.1016/j.isprsjprs.2017.10.008.

Main, R., M. A. Cho, R. Mathieu, M. M. O'Kennedy, A. Ramoelo, and S. Koch. 2011. "An Investigation into Robust Spectral Indices for Leaf Chlorophyll Estimation." *Isprs Journal of Photogrammetry & Remote Sensing* 66 (6): 751–761. https://doi.org/10.1016/j.isprsjprs.2011.08.001.

Minasny, B., B. P. Malone, A. B. McBratney, D. A. Angers, D. Arrouays, A. Chambers, V. Chaplot, et al. 2017. "Soil carbon 4 per mille." *Geoderma* 292:59–86. https://doi.org/10.1016/j.geoderma.2017.01.002.

Mngadi, M., J. Odindi, and O. Mutanga. 2021. "The Utility of Sentinel-2 Spectral Data in Quantifying Above-Ground Carbon Stock in an Urban Reforested Landscape." *Remote Sensing* 13 (21): 4281. https://doi.org/10.3390/rs13214281.

Mutanga, O., and A. K. Skidmore. 2004. "Narrow Band Vegetation Indices Overcome the Saturation Problem in Biomass Estimation." *International Journal of Remote Sensing* 25 (19): 3999–4014. https://doi.org/10.1080/01431160310001654923.

Naidoo, L., M. A. Cho, R. Mathieu, and G. Asner. 2012. "Classification of Savanna Tree Species, in the Greater Kruger National Park Region, by Integrating Hyperspectral and LiDAR Data in a Random Forest Data Mining Environment." *Isprs Journal of Photogrammetry & Remote Sensing* 69:167–179. https://doi.org/10.1016/j.isprsjprs.2012.03.005.

Ndlovu, N., M. Luck-Vogel, B. Schloms, and M. Cho. 2011. "The Quantification of Human Impact on the Dukuduku Indigenous Forest from 1960 to 2008 Using GIS Techniques as a Basis for Sustainable Management." *Fifth Natural Forest and Woodland Symp*.

Odebiri, O., O. Mutanga, and J. Odindi. 2022. "Deep Learning-Based National Scale Soil Organic Carbon Mapping with Sentinel-3 Data." *Geoderma* 411:115695. https://doi.org/10.1016/j.geoderma.2022.115695.

Odebiri, O., O. Mutanga, J. Odindi, and R. Naicker. 2023. "Mapping Soil Organic Carbon Distribution Across South Africa's Major Biomes Using Remote Sensing-Topo-Climatic Covariates and

Concrete Autoencoder-Deep Neural Networks." *Science of the Total Environment* 865:161150. https://doi.org/10.1016/j.scitotenv.2022.161150.

Odebiri, O. O. 2022. *The Application of Deep Learning for Remote Sensing of Soil Organic Carbon Stocks Distribution in South Africa* (Doctoral dissertation).

Proctor, C., B. Lu, and Y. He. 2017. "Determining the Absorption Coefficients of Decay Pigments in Decomposing Monocots." *Remote Sensing of Environment* 199:137–153. https://doi.org/10.1016/j.rse.2017.07.007.

Qabaqaba, M., L. Naidoo, P. Tsele, A. Ramoelo, and M. A. Cho. 2023. "Integrating Random Forest and Synthetic Aperture Radar Improves the Estimation and Monitoring of Woody Cover in Indigenous Forests of South Africa." *Applied Geomatics* 15 (1): 209–225. https://doi.org/10.1007/s12518-023-00497-9.

Sauer, J., V. C. Mariani, L. dos Santos Coelho, M. H. D. M. Ribeiro, and M. Rampazzo. 2022. "Extreme Gradient Boosting Model Based on Improved Jaya Optimizer Applied to Forecasting Energy Consumption in Residential Buildings." *Evolving Systems*: 1–12.

Sibiya, B., R. Lottering, and J. Odindi. 2021. "Discriminating Commercial Forest Species Using Image Texture Computed from a Worldview-2 Pan-Sharpened Image and Partial Least Squares Discriminant Analysis." *Remote Sensing Applications: Society & Environment* 23:100605. https://doi.org/10.1016/j.rsase.2021.100605.

Sims, D. A., and J. A. Gamon. 2002. "Relationships Between Leaf Pigment Content and Spectral Reflectance Across a Wide Range of Species, Leaf Structures and Developmental Stages." *Remote Sensing of Environment* 81 (2–3): 337–354. https://doi.org/10.1016/S0034-4257(02)00010-X.

Song, D., D. Gao, H. Sun, L. Qiao, R. Zhao, W. Tang, and M. Li. 2021. "Chlorophyll Content Estimation Based on Cascade Spectral Optimizations of Interval and Wavelength Characteristics." *Computers and Electronics in Agriculture* 189:106413. https://doi.org/10.1016/j.compag.2021.106413.

Song, S., W. Gong, B. Zhu, and X. Huang. 2011. "Wavelength Selection and Spectral Discrimination for Paddy Rice, with Laboratory Measurements of Hyperspectral Leaf Reflectance." *Isprs Journal of Photogrammetry & Remote Sensing* 66 (5): 672–682. https://doi.org/10.1016/j.isprsjprs.2011.05.002.

Sonobe, R., and Y. Hirono. 2023. "Carotenoid Content Estimation in Tea Leaves Using Noisy Reflectance Data." *Remote Sensing* 15 (17): 4303. https://doi.org/10.3390/rs15174303.

Sonobe, R., and Q. Wang. 2018. "Nondestructive Assessments of Carotenoids Content of Broadleaved Plant Species Using Hyperspectral Indices." *Computers and Electronics in Agriculture* 145:18–26. https://doi.org/10.1016/j.compag.2017.12.022.

Stylinski, C., J. Gamon, and W. Oechel. 2002. "Seasonal Patterns of Reflectance Indices, Carotenoid Pigments and Photosynthesis of Evergreen Chaparral Species." *Oecologia* 131 (3): 366–374. https://doi.org/10.1007/s00442-002-0905-9.

Sun, J., S. Shi, J. Yang, W. Gong, F. Qiu, L. Wang, L. Du, and B. Chen. 2019. "Wavelength Selection of the Multispectral Lidar System for Estimating Leaf Chlorophyll and Water Contents Through the PROSPECT Model." *Agricultural and Forest Meteorology* 266:43–52. https://doi.org/10.1016/j.agrformet.2018.11.035.

Verrelst, J., G. Camps-Valls, J. Muñoz-Marí, J. P. Rivera, F. Veroustraete, J. G. Clevers, and J. Moreno. 2015. "Optical Remote Sensing and the Retrieval of Terrestrial Vegetation Bio-Geophysical Properties–A Review." *Isprs Journal of Photogrammetry & Remote Sensing* 108:273–290.

Verrelst, J., Z. Malenovský, C. Van der Tol, G. Camps-Valls, J. P. Gastellu-Etchegorry, P. Lewis, P. North, and J. Moreno. 2019. "Quantifying Vegetation Biophysical Variables from Imaging Spectroscopy Data: A Review on Retrieval Methods." *Surveys in Geophysics* 40 (3): 589–629. https://doi.org/10.1007/s10712-018-9478-y.

Wan, L., W. Zhou, Y. He, T. C. Wanger, and H. Cen. 2022. "Combining Transfer Learning and Hyperspectral Reflectance Analysis to Assess Leaf Nitrogen Concentration Across Different Plant Species Datasets." *Remote Sensing of Environment* 269:112826. https://doi.org/10.1016/j.rse.2021.112826.

Xu, S., M. Wang, and X. Shi. 2020. "Hyperspectral Imaging for High-Resolution Mapping of Soil Carbon Fractions in Intact Paddy Soil Profiles with Multivariate Techniques and Variable Selection." *Geoderma* 370:114358. https://doi.org/10.1016/j.geoderma.2020.114358.

Yang, C., L. Song, K. Wei, C. Gao, D. Wang, M. Feng, M. Zhang, et al. 2023. "Study on Hyperspectral Monitoring Model of Total Flavonoids and Total Phenols in Tartary Buckwheat Grains." *Foods* 12 (7): 1354. https://doi.org/10.3390/foods12071354.

Yang, J., S. Yang, Y. Zhang, S. Shi, and L. Du. 2021. "Improving Characteristic Band Selection in Leaf Biochemical Property Estimation Considering Interrelations Among Biochemical Parameters Based on the PROSPECT-D Model." *Optics Express* 29 (1): 400–414. https://doi.org/10.1364/OE.414050.

Zhang, J. T., L. Fan, and M. Li. 2012. "Functional Diversity in Plant Communities: Theory and Analysis Methods." *African Journal of Biotechnology* 11 (5): 1014–1022.

Zhang, Y., M. Migliavacca, J. Penuelas, and W. Ju. 2021. "Advances in Hyperspectral Remote Sensing of Vegetation Traits and Functions." *Remote Sensing of Environment* 252:112121. https://doi.org/10.1016/j.rse.2020.112121.

Zhou, X., W. Huang, W. Kong, H. Ye, Y. Dong, and R. Casa. 2017. "Assessment of Leaf Carotenoids Content with a New Carotenoid Index: Development and Validation on Experimental and Model Data." *International Journal of Applied Earth Observation and Geoinformation* 57:24–35. https://doi.org/10.1016/j.jag.2016.12.005.