Smart Agricultural Technology 11 (2025) 100960

Contents lists available at ScienceDirect

Smart Agricultural Technology

journal homepage: www.journals.elsevier.com/smart-agricultural-technology

Leaf area index-based phenotypic assessment of sweet potato varieties using UAV multispectral imagery and a hybrid retrieval approach

Philemon Tsele [a,*], Abel Ramoelo [b,c], Lucy Moleleki [d], Sunette Laurie [e], Whelma Mphela [f], Natasha Tshuma [a]

[a] Department of Geography, Geoinformatics and Meteorology, University of Pretoria, Pretoria 0002, South Africa
[b] Earth Observation Programme, South African National Space Agency, Council for Scientific and Industrial Research (CSIR)'s, Building 10, Meiring Naude, Pretoria 0002, South Africa
[c] Centre for Environmental Studies, Department of Geography, Geoinformatics and Meteorology, University of Pretoria, Pretoria 0002, South Africa
[d] Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria 0002, South Africa
[e] Agricultural Research Council (ARC)-Vegetable, Industrial and Medicinal Plants, Pretoria 0001, South Africa
[f] Innovation Africa, University of Pretoria, Pretoria 0002, South Africa

ARTICLE INFO

Keywords:
Leaf area index (LAI)
UAV imagery
PROSAIL
Sweet potato varieties

ABSTRACT

Phenotyping based on the estimation of plant traits such as the leaf area index (LAI) could aid the identification and monitoring of the sweet potato health, growth status and gross primary productivity. Integrating radiative transfer models (RTMs), active learning algorithms and non-parametric regression methods using unmanned aerial vehicle (UAV) multispectral imagery have the potential for accurately estimating LAI across multiple crop varieties at varying growth stages. This study tested the boosted regression trees (BRT) and kernel ridge regression (KRR) for inversion of the PROSAIL RTM to retrieve LAI across 20 sweet potato varieties during peak growth stage. Furthermore, the study attempted to constrain the inversion process by using active learning (AL) techniques which ensured the selection of informative samples from a pool of RTM simulations. Results show that the most accurate LAI retrieval over the heterogeneous sweet potato canopy was achieved by integrating smaller PROSAIL simulations with the random sampling AL and KRR methods. The LAI retrieval accuracy had a coefficient of determination (R^2) of 0.52, root mean squared error (RMSE) of 0.88 $m^2.m^{-2}$ and relative RMSE of 12.23 %. However, the BRT performance in-comparison to KRR, captured more spatial variability of observed LAI with a better prediction accuracy across the 20 sweet potato varieties. The hybrid approach developed in this study, show potential for accurate phenotyping of LAI dynamics across multiple sweet potato varieties during a matured growth stage. These findings have significant implications for sweet potato breeding programmes that are critical for developing new cultivars in South Africa.

1. Introduction

Sweet potato (*Ipomoea batatas* L.) is an essential crop that has the potential to address malnutrition and guarantee food security in vulnerable populations who are at an increased risk of adverse hunger. Sweet potato has been identified as a crop that is globally underutilized for food security and sustainable rural development, despite its high nutritional value in comparison to staple crops such as maize and wheat [41,53]. In particular, smallholder farmers play a critical role in the establishment of formal and informal markets which are accessible to consumers [34,41,53]. This potentially leads to an increase in the consumption of sweet potato and thus necessitates higher production. Precision agriculture will ensure that productivity is increased in an efficient and effective way.

The estimation of vegetation biophysical variables such as leaf area index (LAI) and leaf chlorophyll concentration (LCC) could assist farmers in assessing and monitoring crops at different stages [35,61]. In particular, the LAI, defined as the one-sided leaf area per unit of horizontal surface area Jonckheere et al. [29] is an important indicator of plant canopy structure and growth, and also forms an essential input in climate models to determine ecosystem productivity. For example, studies have shown that LAI can be used as a proxy for understanding the morphological and/or physiological characteristics of crop cultivars at different growth stages [7]. Another study found LAI to be a key

* Corresponding author.
E-mail address: philemon.tsele@up.ac.za (P. Tsele).

https://doi.org/10.1016/j.atech.2025.100960
Received 18 March 2025; Received in revised form 12 April 2025; Accepted 18 April 2025
Available online 18 April 2025

parameter in explaining how the variation in soil moisture and water depth influence the growth performance of sweet potato [33]. Furthermore, LAI is a good indicator of above-ground growth which reportedly has a direct influence in crop yield below the ground [32]. Accurate measurements of LAI can provide an indirect measure of the crop nutrient status in different growth stages and environments.

Unmanned aerial vehicle (UAV) remote sensing provides an alternative approach to destructive, expensive and time-consuming field campaigns. This approach can provide imagery at very high (cm-level) spatial resolution and flexible temporal resolutions, making it ideal for mapping and monitoring crop biophysical variables such as LAI, particularly in breeding and field trial plots [35]. The estimation of vegetation biophysical variables from remote sensing data, is carried out using three approaches, namely the parametric, non-parametric and radiative transfer models (RTMs). Parametric and non-parametric statistical approaches are not always robust and require a lot of in situ data, but site, season and data specific. Compared to the two former empirical approaches, RTMs have minimum reliance on *in-situ* data in that, they use the physical laws to accurately describe the spectral variation of canopy reflectance as a function of viewing and illumination geometry, leaf, canopy and soil background characteristics [25].

Numerous studies have explored the integrated use of RTMs with machine learning algorithms on UAV imagery for the estimation of crop biophysical variables [6,18,28,35]. However, these studies mainly focused on crops such as maize, wheat, sugar beet, potato and sunflower, with each crop having either a single cultivar or very few cultivars (varieties). There are fewer emerging studies based on empirical or statistical modelling approaches using LAI as a proxy for understanding the morphological and/or physiological characteristics of the sweet potato varieties at different growth stages [33,46]. Furthermore, there is a notable steady increase in studies that use empirical approaches to model sweet potato yield estimates and also to explore how external variables affect the shape and size properties of sweet potato, as these are important in determining grade and monetary value [1,3,32,36,54,55]. Based on reviewed literature thus far, the current study is the first to explore the integrated use of RTMs with machine learning and active learning (AL) algorithms on UAV imagery for the estimation of LAI over a heterogeneous sweet potato canopy comprising 20 varieties.

A major challenge with using RTM data is that it may contain redundant or non-diverse samples which do not improve on the prediction accuracy of the resulting regression model [59]. One of the ways in addressing this limitation is through the use of AL sample selection algorithms to: (i) disregard the non-diverse samples from the pool of RTM-simulated reflectance samples, and (ii) optimise the simulated training dataset to contain only intelligent or informative samples needed for improving the regression model's retrieval accuracy [43].

The aim of this study was to explore the hybrid model of integrating PROSAIL simulated reflectance samples with two non-parametric regression methods using UAV multispectral imagery, to estimate LAI over 20 sweet potato varieties at a matured growth stage. The specific objectives of this study were to: (i) generate different PROSAIL databases containing varying simulated canopy reflectance (training) samples to determine their influence on model estimation accuracy (ii) apply several AL sample selection algorithms to eliminate redundant samples from each database, and (iii) compare the performance of kernel ridge regression (KRR) and boosted regression trees (BRT) for the retrieval of LAI. This study has significant implications for: (i) sweet potato breeding programmes critical for developing new cultivars in South Africa and (ii) small to large scale farmers in obtaining accurate maps of sweet potato LAI, which are essential for assessing and monitoring crop growth at different stages.

2. Study area

The study site is based on a small trial plot covering an area of about 500 m^2 located between 28°15′36.07″E, 25°44′59.43″S and 28°15′37.39″E, 25°44′59.83"S at the experimental farm of the University of Pretoria, South Africa (Fig. 1). The Roodeplaat Vegetable, Industrial and Medicinal Plants (RVIMP) institute of the Agricultural Research Council (ARC) in South Africa has been involved in sweet potato breeding for >70 years, and currently 33 sweet potato cultivars have been developed in South Africa. In particular, sweet potato breeding programme at RVIMP-ARC aims to evaluate a new set of sweet potato varieties for stress adaptation, growth performance and productivity. This evaluation exercise is critical for developing new sweet potato cultivars in South Africa.

In this study, there were 20 unique sweet potato varieties and three replicates of each variety, comprising a total of 20 × 3 = 60 varieties planted in the field (Table 1). The geographic coordinates of each plot linked to specific varieties shown in yellow dots in Fig. 1(C) were recorded using the Global Navigation Satellite System – Real Time Kinematic (GNSS-RTK) method [49].

Furthermore, the study site (Fig. 1) is characterised by a relatively flat terrain and falls within the summer rainfall region of South Africa. During the period of the sweet potato trials from planting date (28 January 2024) until peak growth stage (in late May 2024), the study site received monthly average rainfall in the range 175 mm – 10 mm coupled with monthly average temperature in the range 21.5 °C – 18.5 °C (Fig. 2). The observed decline in rainfall and temperature toward peak growth stage is due to seasonal changes from summer to the beginning of winter. The rainfall and temperature estimates shown in Fig. 2 were obtained from the Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) dataset [45] and ERA5-Land dataset [42], respectively.

3. Material and methods

3.1. Data used

3.1.1. Description of LAI field measurements

Since planting on 28 January 2024, the LAI and LCC field measurements of the sweet potato varieties were taken from 24 February to 27 May 2024 at approximately 30-day intervals spanning early growing to the matured stages (Table 2). In particular, the LAI measurements were taken using the LI-COR LAI-2200C Plant Canopy Analyzer [10] under clear skies from late morning hours at about 11:00 until early afternoon at around 13:00 in-order to minimize variations of the sun zenith angle in all the plots. A 180° view cap was applied to the LAI-2200C to block the azimuthal field of view of the operator. Each plot had a dimension of 2.1 m x 6 m and thus, the 20 unique sweet potato varieties with three replicates of each (comprising a total of 20 × 3 = 60 varieties) were distributed across 60 plots.

In each plot, the LAI-2200C instrument was used to record the incident photosynthetically active radiation (PAR; 400–700 nm) above the sweet potato canopy, followed by multiple recordings of the below sweet potato canopy PAR in-order to calculate the average LAI per plot (Table 2). Furthermore, LCC field measurements of 10 randomly selected leaves of the sweet potato were taken using the SPAD 502 Plus chlorophyll meter, and averaged into single value representing the LCC per plot. The SPAD unitless LCC values across the 60 plots were converted into $\mu g.cm^{-2}$ by applying an empirical calibration method described in Cerovic et al. [4]. In this study, both LAI and LCC measurements for the peak growth stage (27 May 2024) were used for parameterising the models and evaluating their performance.

3.1.2. UAV imagery and pre-processing

The DJI Matrice 300 RTK drone mounted with a Micasense RedEdge-P camera and Downwelling Light Sensor 2 (DLS-2) was used to acquire multispectral imagery at nadir of the sweet potato field on 24 February, 29 March, 24 April and 27 May 2024 (Fig. 3). The Micasense RedEdge-P camera has five multispectral bands (i.e. blue 475 nm, green 560 nm, red 668 nm, red-edge 717 nm and near-infrared 842 nm) and one

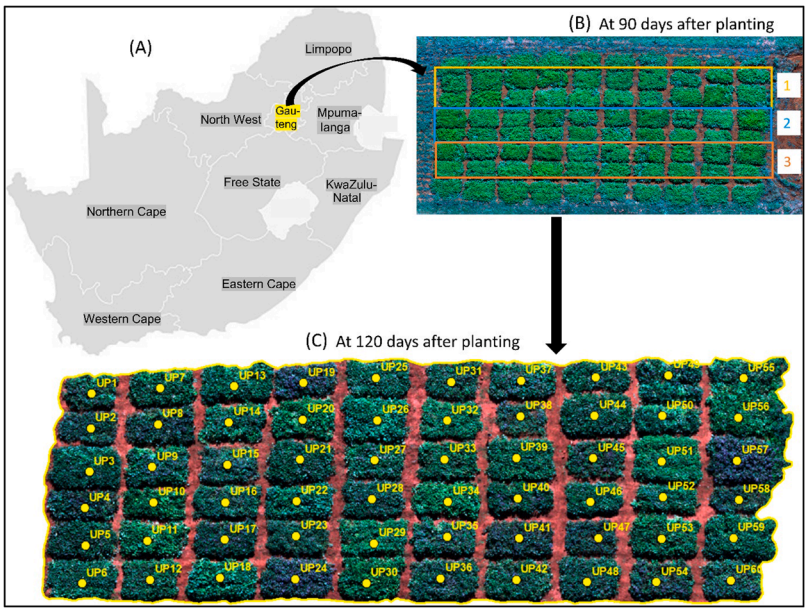

Fig. 1. Study area map showing the South African provincial boundaries (A), a true colour drone composite image of the 20 sweet potato varieties per highlighted row with three replicates, captured at 90 days after planting (B), and a true colour drone composite image captured at matured stage (120 days after planting). The yellow labels or dots indicate centre locations of all varieties recorded by use of a high precision GNSS system. The labels shown on the image (C) represents the varieties shown in Table 1.

Table 1
GPS labels in Fig. 1(C) and their corresponding sweet potato variety code.

No.	Field Label	Variety
1	UP 1, 45, 54	NC51–6
2	UP 2, 47, 58	Evangeline
3	UP 3, 23, 50	199062.1
4	UP 4, 8, 41	2004–9–2 (ARC-SP-2)
5	UP 5, 21, 44	NC55–2
6	UP 6, 25, 46	Blesbok
7	UP 7, 33, 53	Bophelo
8	UP 9, 35, 37	NC-–51–15 (ARC-SP-12)
9	UP 10, 30, 56	Ndou
10	UP 11, 20, 34	Monate
11	UP 12, 14, 52	NC51–1 (ARC-SP-11)
12	UP 13, 39, 59	2008–12–4 (ARC-SP-5)
13	UP 15, 38, 60	Beauregard
14	UP 16, 36, 55	2014–13–1 (ARC-SP-6)
15	UP 17, 40, 49	2014–7–3 (ARC-SP-7)
16	UP 18, 22, 26	2014–14–5 (ARC-SP-8)
17	UP 19, 24, 57	2010–15–2 (ARC-SP-10)
18	UP 27, 42, 43	Khumo
19	UP 29, 32, 51	FS10–25
20	UP 28, 31, 48	2015–2–1

panchromatic band (not used in this study). The camera has a resolution of 1456 × 1088 (i.e. 1.58 megapixels per multispectral band) coupled with a horizontal and vertical field of view of 49.6° x 38.3° respectively. In this study, it was flown at an altitude of 30 m above ground level (AGL) covering a ground sampling distance (GSD) of 1.9 cm per pixel. The AGL used and GSD achieved, were considered to be ideal for capturing the finer details of the sweet potato growth stages i.e. from seedling, vegetative and all the way to matured stages.

Prior every flight, the UAV was connected to the DJI d-RTK 2 Mobile Station (https://www.dji.com/global/d-rtk-2/info) which provides the UAV with real-time differential corrections for generating centimeter-level positioning data. The UAV flights were conducted on clear-sky days, and strictly between 11:30 a.m. and 12:30 pm. which was considered the ideal time when there is minimal shadow effect and optimal solar irradiance. Images of the MicaSense Calibrated Reflectance Panel (CRP) were taken using the multispectral camera before and after every UAV flight. These images were later used in Agisoft Metashape software (version 2.0.0) to calibrate the reflectance of the multispectral data (Fig. 4). These steps are consistent with the Agisoft Metashape image calibration method that uses a single calibration panel such as the CRP [9]. It has been reported in Daniels et al. [9] that the image calibration method embedded in commercial software packages such as Agisoft and Pix4D Fields, has easier or less complicated processes to convert raw image into usable reflectance maps with sufficient accuracy (particularly in clear-sky conditions) comparable to other sophisticated calibration methods such as the empirical line method using multiple reference targets [39].

The mean spectral reflectance of the 20 sweet potato varieties was computed from the drone reflectance image (Fig. 4). Each spectral curve represents the average spectra from the three replicates of each sweet potato variety. The spectral profile is more characteristic of a greener and healthy sweet potato canopy. Although, the observed spectral variability may be an indication of the heterogeneity of the sweet potato canopy across the plots, due to for example, different varieties coupled with their associated biophysical (leaf structure, area, coverage and chlorophyll) and biochemical (leaf water and nitrogen content) properties at a leaf-level.

3.2. Methods used

3.2.1. Schematic workflow

Fig. 5 show a schematic workflow summarizing the various phases of the methodology that were implemented in this study. These phases are discussed in subsequent sections of the paper.

3.2.2. PROSAIL model parameterization and simulated spectra

The PROSPECT and SAIL (hereafter PROSAIL) RTM models [25] are available to the public in a toolbox named Automated Radiative Transfer Models Operator (ARTMO) (https://artmotoolbox.com/). The toolbox was developed and runs in Matlab. In this study, PROSAIL in ARTMO was used to generate databases of synthetic (simulated) reflectance samples of the sweet potato varieties, based on a combination of adopted and site-

During parameterisation, the range values for LCC and LAI were

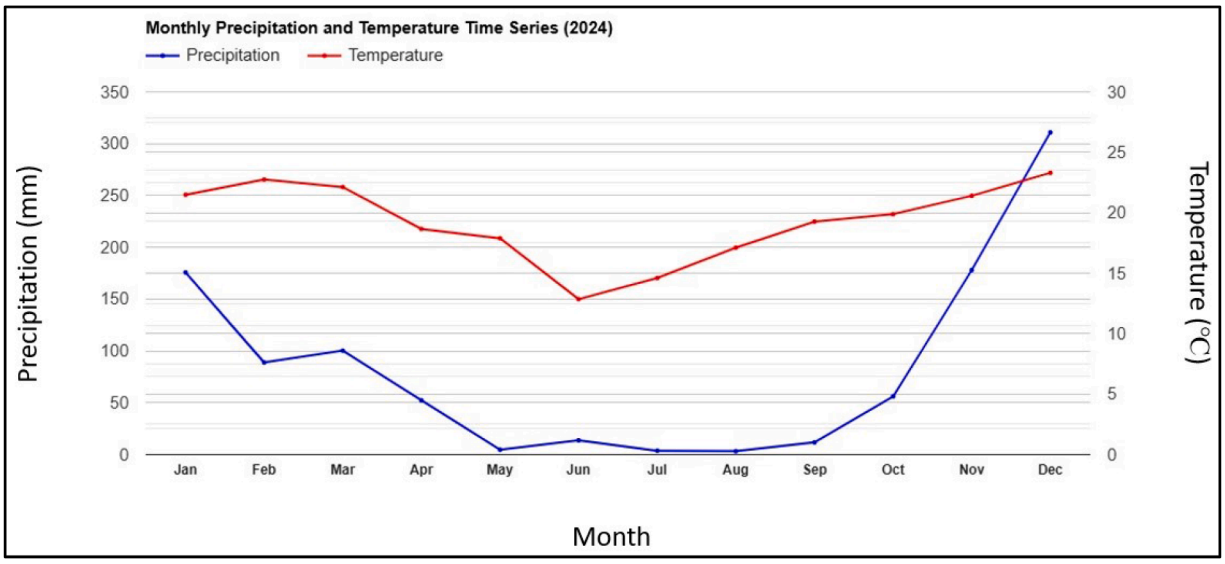

Fig. 2. Monthly average precipitation and temperature at the experimental farm of the University of Pretoria for the period 2024.

Table 2
Summary statistics of the field-measured biophysical variables of LCC ($\mu g.cm^{-2}$) and LAI ($m^2.m^{-2}$) of 20 sweet potato varieties planted in 60 plots or compartments within the field. StDev denotes the standard deviation.

Date	Variable	Samples	Min.	Max.	Mean	Median	StDev
24 February 2024	LCC	60	36.64	75.07	52.74	51.70	8.74
	LAI	60	1.52	4.11	2.62	2.53	0.53
26 March2024	LCC	60	48.12	90.57	62.79	60.82	9.02
	LAI	60	4.68	9.16	7.22	7.24	1.04
24 April 2024	LCC	60	45.44	97.36	63.43	62.91	10.13
	LAI	60	6.25	11.30	8.58	8.38	1.07
27 May 2024	**LCC**	**60**	**42.85**	**125.15**	**59.90**	**56.63**	**14.75**
	LAI	**60**	**4.31**	**9.60**	**7.22**	**7.51**	**1.27**

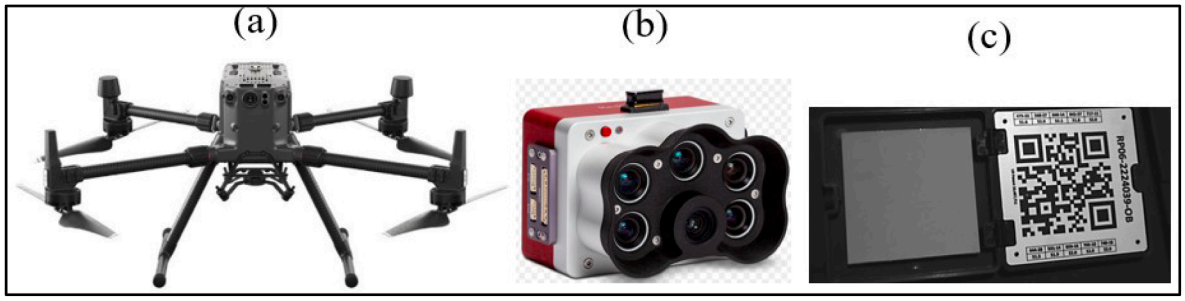

Fig. 3. The equipment used to acquire UAV multispectral imagery i.e. (a) DJI Matrice 300 UAV, (b) Micasense RedEdge-P camera, and (c) MicaSense Calibrated Reflectance Panel.

based on actual field measurements; whereas the range values for N and ratio of diffuse to downward irradiance were based on literature [24, 56]. In this study, for the leaf structure parameter N, given that sweet potato is a dicotyledonous plant and also that the sweet potato leaves were non-senescent (at 120 days), we relied on the study by Jacquemoud and Baret [24] that recommended N values between 1.5 and 2.5 for a plant that has a dicotyledon leaf type. The dry matter content was measured in the lab by considering the difference between the original and dry weights of the sweet potato leaves. The EWT was estimated using the expression in Féret et al. [19] which computes the difference between leaf fresh mass (FW) and leaf dry mass (DW) divided by the leaf area (A). EWT was estimated based on measurements obtained from nine randomly selected sweet potato varieties (namely, Evangeline, NC55–2, Blesbok, Bophelo, Ndou, Monate, Beauregard, FS10–25 and Khumo (Table 1)) and it was found that the EWT ranged from 0.023 to 0.044 g/cm^2 with an average of 0.03 g/cm^2. Hence, a fixed EWT value of 0.03 g/cm^2 was used for parameterisation (Table 3).

Furthermore, the hot-spot parameter, which represents the peak backscatter when the sun is directly behind the sensor [15,31] was computed by taking the ratio of field-measured sweet potato leaf area (in cm) to the canopy height (in cm) per plot and variety. The soil brightness coefficient was estimated from the UAV multispectral bands i.e. red and green bands, using the soil brightness index equation described in Marques et al. [40]. The sun-sensor geometrical parameters such as the solar zenith and sensor observation angles were determined during UAV flight time.

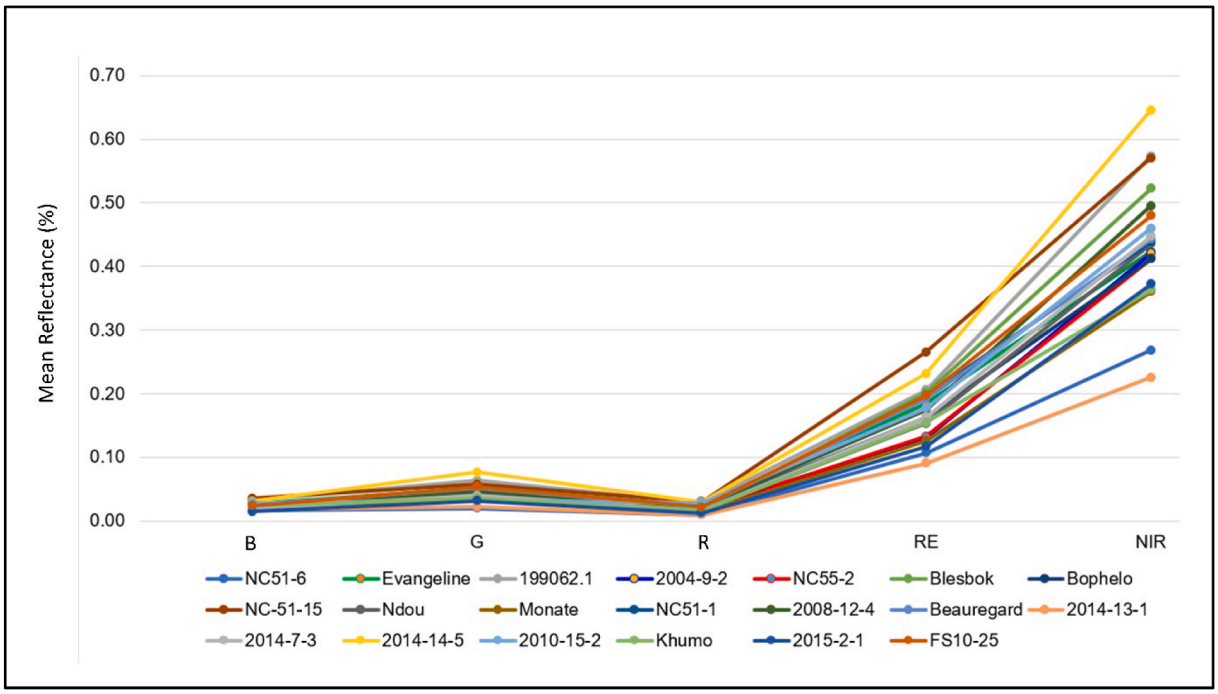

Fig. 4. Mean spectral reflectance of the 20 sweet potato varieties at matured stage extracted from the 27 May 2024 UAV image (at 120 days).

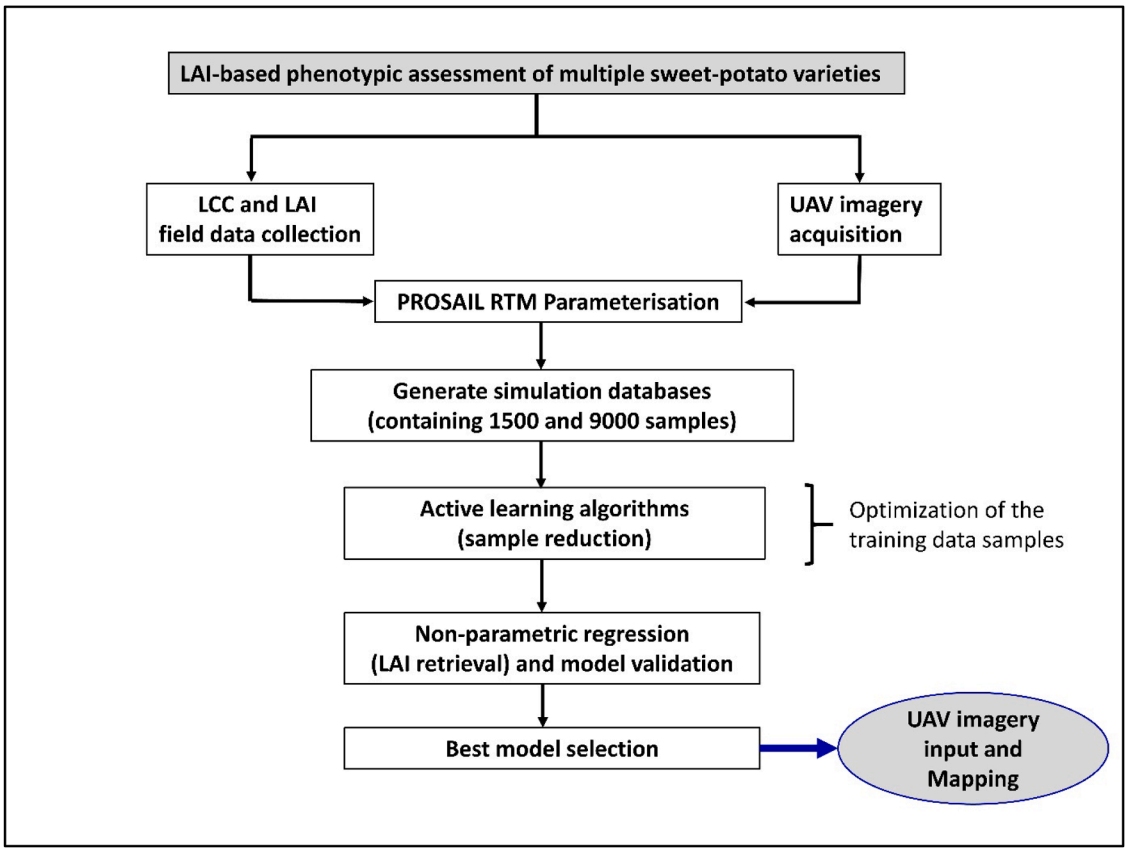

Fig. 5. A hybrid retrieval workflow showing the various phases of the methodology that were implemented in this study.

3.2.3. PROSAIL simulated spectra

The sensor settings related to the Micasense RedEdge-P with five bands that range from 475 nm to 842 nm (centre wavelength) were chosen. PROSAIL-simulated data samples were generated and stored in a lookup table (LUT) database in ARTMO. Studies have explored working with a variable range of simulation samples from as little as 250 up to 50,000 in-order to find optimal samples that can be used for training the regression models to achieve a better retrieval performance [35,57]. It

Table 3
PROSAIL model parameterisation. Ave: average or mean, StDev: standard deviation.

Model parameters	Unit	Range	Distribution	Source
Leaf parameters: PROSPECT-4 model				
Leaf chlorophyll content (LCC)	[μg/cm^2]	42.85 −125.15	Gaussian (Ave: 59.90; StDev: 14.75)	Field data
Leaf structure (N)	Dimensionless	1.5 – 2.5	Uniform	Jacquemoud and Baret [24]
Equivalent water thickness (EWT)	[g/cm^2]	0.03	Fixed	Field data
Dry matter	[g/cm^2]	0.0062 - 0.0145	Uniform	Field data
Canopy parameters: 4SAIL model				
Leaf area index (LAI)	[m^2/m^2]	4.31 - 9.60	Gaussian (Ave: 7.22; StDev: 1.27)	Field data
Average leaf angle (ALA)	[°]	50 - 80	Uniform	Observation during field data collection
Hot-spot effect	[fraction]	0.26 - 0.48	Gaussian (Ave: 0.36; StDev: 0.05	Field data
Ratio of diffuse to downward irradiance	[fraction]	0.07 – 0.22	Uniform	Tongwane et al. [56]
Soil brightness	Dimensionless	0.50	Fixed	Soil brightness index on UAV bands [40]
Solar zenith angle	[°]	0	Fixed	Field observation during flight time
Sensor zenith angle	[°]	0	Fixed	UAV Gimble angle setting

was reported that, working with a large number of simulation samples (e.g. of 100,000 or more), potentially creates largely-redundant samples for regression [59].

In this study, PROSAIL-simulated data were generated containing a pool of small (1500) to large (9000) samples of synthetic canopy reflectance, in-order to assess their influence on LAI retrieval performance. This generated pool of simulated data was used for training the non-parametric regression methods (NPRMs) for retrieving the sweet potato LAI. Furthermore, there was an attempt to constrain the inversion process by using AL techniques (explained later) which ensured the selection of only the best possible samples from a pool of RTM-simulations for use by the NPRMs.

3.2.4. Non-parametric regression methods (NPRM)

This study evaluated two nonlinear NPRMs, widely used in the literature for estimating vegetation biophysical variables [58,60] namely: kernel ridge regression (KRR) and boosted regression trees (BRT). These methods were chosen in this study because: (i) of their ability to handle nonlinear relationships or high-dimensional data, (ii) they define the regression functions based on input data, and (iii) they can optimise the regression models by learning the training data [60].

The KRR is a supervised learning model that makes use of kernel functions for data analysis and pattern identification [22]. KRR is a family of the Least Squares Support Vector Machine Classifiers [52] which maps the training samples into a higher dimensional feature space and builds a regression function which represents a nonlinear regression in the original input space [48]. An optimal function would minimize the squared residuals and lead to improved biophysical variable retrieval. KRR in ARTMO software package required the tuning of the kernel function, regularization and optimization parameters. The optimization was carried out using the standard cross validation procedure.

On the other hand, BRT is a machine learning technique that combines decision trees with boosting methods, which helps to improve model accuracy [12]). Unlike random forests, which treat all data points equally when creating trees [2], BRT give more importance to data points that were previously mis-predicted, allowing the model to continuously learn and improve. This method is particularly useful for predicting environmental patterns, like species distributions or climate effects, because it accounts for complex interactions between variables [20]. The model's accuracy depends on two key settings: tree complexity (tc), which determines how detailed each tree is, and learning rate (lr), which controls how much influence each new tree has on the overall prediction. A well-tuned BRT model usually requires at least 1000 trees for reliable predictions. It is especially useful when working with smaller datasets, as it can adjust and refine predictions more effectively over time.

3.2.5. Active learning techniques

AL techniques use selection criterion algorithms [38] to select informative samples from a large synthetic training database in-order to improve the model's estimation accuracy [43]. The AL techniques used in this study falls under two main categories namely, uncertainty criteria algorithms and diversity criteria algorithms. The former, uses variance-based algorithms [16] to select from a large pool of samples, only the those with the least confidence. Whereas, the latter uses a variety of distance-related metrics [14,17,44] to select the most diverse samples and thereby disregarding the redundant samples from a pool of RTM-simulated reflectance samples. In particular, the uncertainty criteria algorithm used in this study was the Residual active learning (RSAL: [16]). RSAL applies the residual model to estimate the prediction error linked to each obtained prediction and ranks the different predictions according to their estimated residual errors. A selection of the samples related to the predictions with the highest prediction errors is performed and are therefore, considered to be the most uncertain samples that will not be considered in optimal final training set [59].

Furthermore, the diversity criteria algorithms used in this study were the Angle-based diversity (ABD: [14]), Euclidean distance-based diversity (EBD: Douak et al. [17]) and Cluster-based diversity (CDB: [44]). The ABD algorithm measures the degree of diversity between samples in the initial training set (i.e. subset of n random samples) and those in the RTM-simulated database using the cosine angle distance. The samples with smallest cosine angles are ranked low because they represent samples (also referred to as reflectance-variable pairs) that are redundant and similar to those already accounted for in the training set. However, samples with the largest cosine angles are ranked high and added to the training set until it become optimal [8]. The EBD algorithm works similarly to ABD, however the difference is that EBD measures the degree of diversity by calculating the Euclidean distances [17] between the samples in the initial training set and those in the RTM-simulated database. Samples with the farthest distance are ranked high and added to the training set until it become optimal.

Another diversity criteria algorithm used was the CDB, which is a standard cluster based technique that applies the k-means clustering algorithm [26] to partition the initial training set into a series of labelled n clusters in the feature space. The number of clusters n, is set to the number of samples to add in each iteration of the algorithm [59]. The cluster centroid is determined for each cluster and thereafter, iteratively selects the nearest sample (from a pool of unlabelled synthetic samples) to the cluster centroid. Generally, samples within the same cluster are correlated and (in this case) characterised by minimal variable variations that might produce virtually similar spectra. Therefore, the most informative samples within the clusters would be selected either based on their distribution and/ or level of uncertainty [14]. An improved version of the CDB algorithm couples the diversity measure with uncertainty analysis of the samples [44].

Last but not least, the Random sampling (RS) AL algorithm was used in this study. RS falls under diversity criteria algorithms and it is considered the most straightforward algorithm in that it gives every sample in the RTM-simulated spectra database equal probability of being selected. Basically, RS selects at random, a pre-defined number of samples within a large pool of unlabelled RTM-simulated spectra, and add them to the training set in-order to obtain an optimised training set

3.2.6. Evaluation of model prediction accuracies and performance of the AL algorithms

In this study, 60 field measurements of LAI were used for validation. In particular, the standard cross-validation method [51] was used to evaluate the retrieval performance of the NPRMs. In addition, cross-validation was used to evaluate the relative performance of the AL algorithms when applied to the NPRMs. During cross-validation in ARTMO, the LAI field dataset was randomly divided into $k = 10$ equal-sized sub-datasets. We defined five iterative validation steps and, in each step, the k sub-datasets were used only once as a validation dataset for model testing. The hybrid retrieval performance of the NPRMs (with the integration of AL algorithms) was evaluated using statistical performance metrics such as the coefficient of determination (R^2), root mean-squared error (RMSE), and relative root mean-squared error (RRMSE). These metrics are widely used in numerous studies involving the estimation of vegetation biophysical and/or biochemical parameters, for example Li et al. [35], Chakhvashvili et al. [6], Jay et al. [28], Duan et al. [18] and Chen et al. [7].

$$R^2 = 1 - \frac{\sum \left(e_k^N - \overline{e}_k\right)^2}{\sum \left(e_k - \overline{e}_k\right)^2} \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{k=1}^{N} (e_k - m_k)^2}{n}} \quad (2)$$

$$RRMSE = \frac{RMSE}{\overline{m}_k} \times 100 \quad (3)$$

where m_k is the observed biophysical variable i.e. LAI and e_k is the model predicted biophysical variable i.e. LAI, $\overline{m}_k$, and $\overline{e}_k$ denotes the respective means of observed and model predicted biophysical variables, n is the sample size, and N is the number of errors.

The R^2 shown in Eq. (1) was computed for each model to measure the goodness of fit. This was followed by the computation of RMSE shown in Eq. (2) which indicate the amount of error expressed in the units of the biophysical variable of interest i.e. $m^2.m^{-2}$ for LAI. RMSE can range from 0 to ∞ and a lower value (closer to 0), indicate an accurate model [5]. Additionally, the RRMSE shown in Eq. (3) was used to facilitate comparison of model accuracies, where model accuracy was regarded as either excellent (RRMSE<10 %), good (10 %<RRMSE<20 %), fair (20 %<RRMSE<30 %) or inadequate (RRMSE>30 %) [23,27,47].

4. Results

4.1. Analysis of the LAI and LCC field measurements

The field measurements across the 60 plots (Table 2), resembled an approximately normal distribution, which was inferred from the proximity of the respective mean and median values per variable. The standard deviations of the LAI and LCC show minimal increasing variability of the sweet potato canopy from early to matured growth stage. The heterogeneity of the sweet potato varieties and the canopy in general, increased with different growing stages. The trend observed in the respective mean and range values of LAI and LCC suggested a positive phonological development of the sweet potato varieties. The current study focused on the peak growth stage, and thus the data used in this study was that of 27 May 2024 (Table 2). The aforementioned variability of field measurements is important when parameterising the PROSAIL RTM to produce synthetic reflectance samples that capture a broader range of the sweet potato conditions across the plots.

4.2. KRR retrieval performance of sweet potato LAI

The estimation of sweet potato LAI during the matured stage (in 120 days) using KRR integrated with AL techniques showed varying accuracies between the different RTM simulation samples (Table 4 and Table 5). In particular, the most accurate retrievals of LAI over the heterogeneous sweet potato canopy were achieved using a smaller pool (1500) PROSAIL simulations (Table 4) compared to the 9000 simulation samples (Table 5). Fig. 6 show that, each of the AL algorithms started with an initial training set of 15 random samples taken from a database of 1500 PROSAIL-RTM samples, and through numerous iterations, grew this set by adding informative (also referred to as smart or intelligent) samples until it became optimal with approximately 90 training samples.

The optimal samples led to adequate sweet potato LAI retrieval performance by KRR when integrated with the RS AL method. When KRR is used with optimised samples from the RS AL technique, the LAI retrieval accuracy reached an R^2 of 0.52 and a RMSE of 0.88 $m^2.m^{-2}$. Additionally, the model prediction error was the lowest with a RRMSE of 12.23 % signalling a good model accuracy. The second best retrieval accuracy was achieved again using smaller (1500) PROSAIL simulations when KRR is used with optimised samples from the EBD AL technique i.e. R^2 of 0.45 and RRMSE of 12.95 % (Table 4).

Optimised samples derived from other AL techniques (such as the ABD, CBD and RSAL) applied to the 1500 PROSAIL simulations, yielded deteriorating LAI retrieval accuracies with R^2 values ranging from 0.18 to 0.22 and RRMSE values between 15.47 % and 16.74 %. A virtually similar trend is observed when LAI retrievals were attempted on 9000 PROSAIL simulations (Table 5) thus signalling that, the integration of larger synthetic reflectance samples with KRR applied to UAV imagery may not provide an accurate LAI-based phenotypic assessment of multiple sweet potato varieties at plot-scale.

4.3. Boosted regression trees retrieval performance of sweet potato LAI

The estimation of sweet potato LAI during the matured stage (in 120 days) using BRT integrated with AL techniques showed varying accuracies between the different RTM simulation samples (Table 6 and Table 7). While the BRT LAI estimation accuracies were generally lower compared to those associated with KRR, the results show that the most accurate retrievals of LAI over the heterogeneous sweet potato canopy based on BRT was achieved using smaller (1500) PROSAIL simulations (Table 6) compared to the larger i.e. 9000 simulation samples (Table 7). Fig. 7 shows that, each of the AL algorithms started with an initial training set of 15 random samples taken from a database of 1500 PROSAIL-RTM samples, and through numerous iterations, grew this set by adding informative samples until it became optimal, with training samples ranging between 25 and 32. The optimization by AL techniques followed by the retrieval process using the BRT (or KRR) method

Table 4
The leaf area index (LAI) retrieval performance of kernel ridge regression (KRR) based on optimised training samples selected from a database of 1500 synthetic reflectance samples, using various active learning (AL) algorithms i.e. Angle-based diversity (ABD), Cluster-based diversity (CDB), Euclidean distance-based diversity (EBD), Random sampling (RS) and Residual active learning (RSAL).

Algorithm	RMSE ($m^2.m^{-2}$)	RRMSE (%)	R^2
Angle Based Diversity [ABD]	1.21	16.74	0.18
Clustering-Based Diversity [CBD]	1.12	15.47	0.22
Euclidean Diversity [EBD]	0.93	12.95	0.45
Random Sampling [RS]	**0.88**	**12.23**	**0.52**
Residual Active Learning [RSAL]	1.14	15.86	0.20

Table 5
The leaf area index (LAI) retrieval performance of kernel ridge regression (KRR) based on optimised training samples selected from a database of 9000 synthetic reflectance samples, using various active learning (AL) algorithms i.e. Angle-based diversity (ABD), Cluster-based diversity (CDB), Euclidean distance-based diversity (EBD), Random sampling (RS) and Residual active learning (RSAL).

Algorithm	RMSE ($m^2.m^{-2}$)	RRMSE (%)	R^2
Angle Based Diversity [ABD]	1.26	17.46	0.05
Euclidean Diversity [EBD]	2.71	37.53	0.10
Clustering-Based Diversity [CBD]	1.13	15.63	0.22
Random Sampling [RS]	1.13	15.63	0.20
Residual Active Learning [RSAL]	1.07	14.78	0.32

happened simultaneously.

The optimal samples led to a modest retrieval performance of the sweet potato LAI by BRT when integrated with the ABD AL method (Table 6 and Fig. 7). When the BRT is used with optimised samples from the ABD AL technique, the LAI retrieval accuracy reached an R^2 of 0.50 and a RMSE of 0.89 $m^2.m^{-2}$. Additionally, the model prediction error was lower with a RRMSE of 12.34 % signalling a good model accuracy. The second best retrieval accuracy by BRT was again achieved using a smaller PROSAIL database with 1500 simulation samples, where BRT was used with optimised samples from the CBD AL technique i.e. R^2 of 0.28 and RRMSE of 15.27 % (Table 6).

Overall, the sweet potato LAI retrieval by BRT based on the optimised samples derived from other AL techniques (such as the CBD, EBD, RS and RSAL) applied to the 1500 PROSAIL simulations, yielded low LAI retrieval accuracies with R^2 values ranging from 0.19 - 0.28 and RRMSE values between 15.27 % - 16.05 % (Table 6). The sweet potato LAI retrieval by BRT showed further deteriorating accuracies (with R^2 values ranging between 0.11 and 0.26 and RRMSEs' in the range 15.84 % to 18.03 %) when LAI retrievals were attempted on 9000 PROSAIL simulations (Table 7) thus signalling that, the integration of larger synthetic reflectance samples with BRT (similar to KRR) applied to UAV imagery may not provide an accurate LAI-based phenotypic assessment of multiple sweet potato varieties at plot-scale. These findings suggest that, working with a large database of synthetic reflectance samples (e.g. of 9000 or possibly more), potentially creates largely-redundant samples for modelling the LAI of multiple sweet potato varieties in the current study. However, using lesser PROSAIL simulation samples showed a better LAI retrieval performance in-terms of explained variability (R^2) and model prediction error (RMSE and RRMSE) over 20 sweet-potato varieties and their replicates in the current study.

4.4. KRR spatial prediction of sweet-potato LAI

Fig. 8 shows the prediction map of sweet potato LAI, using the best-performing kernel ridge regression (KRR) model integrated with the Random sampling (RS) active learning (AL) method i.e. R^2 of 0.52, RMSE of 0.88 $m^2.m^{-2}$ and RRMSE of 12.34 % (Table 4). In particular, the KRR model spatially predicted LAI values in the range of 0.20 – 12.8 $m^2.m^{-2}$ across the 20 sweet potato varieties at matured growth stage. This

Table 6
The leaf area index (LAI) retrieval performance of the boosted regression trees (BRT) based on optimised training samples selected from a database of 1500 synthetic reflectance samples, using various active learning (AL) algorithms i.e. Angle-based diversity (ABD), Cluster-based diversity (CDB), Euclidean distance-based diversity (EBD), Random sampling (RS) and Residual active learning (RSAL).

Algorithm	RMSE ($m^2.m^{-2}$)	RRMSE (%)	R^2
Angle Based Diversity [ABD]	**0.89**	**12.34**	**0.50**
Clustering-Based Diversity [CBD]	1.10	15.27	0.28
Euclidean Diversity [EBD]	1.12	15.46	0.28
Random Sampling [RS]	1.15	15.98	0.19
Residual Active Learning [RSAL]	1.16	16.05	0.20

Table 7
The leaf area index (LAI) retrieval performance of the boosted regression trees (BRT) based on optimised training samples selected from a database of 9000 synthetic reflectance samples, using various active learning (AL) algorithms i.e. Angle-based diversity (ABD), Cluster-based diversity (CDB), Euclidean distance-based diversity (EBD), Random sampling (RS) and Residual active learning (RSAL).

Algorithm	RMSE ($m^2.m^{-2}$)	RRMSE (%)	R^2
Angle Based Diversity [ABD]	1.16	16.13	0.18
Clustering-Based Diversity [CBD]	1.28	17.71	0.17
Euclidean Diversity [EBD]	1.17	16.18	0.18
Random Sampling [RS]	1.30	18.03	0.11
Residual Active Learning [RSAL]	1.14	15.84	0.26

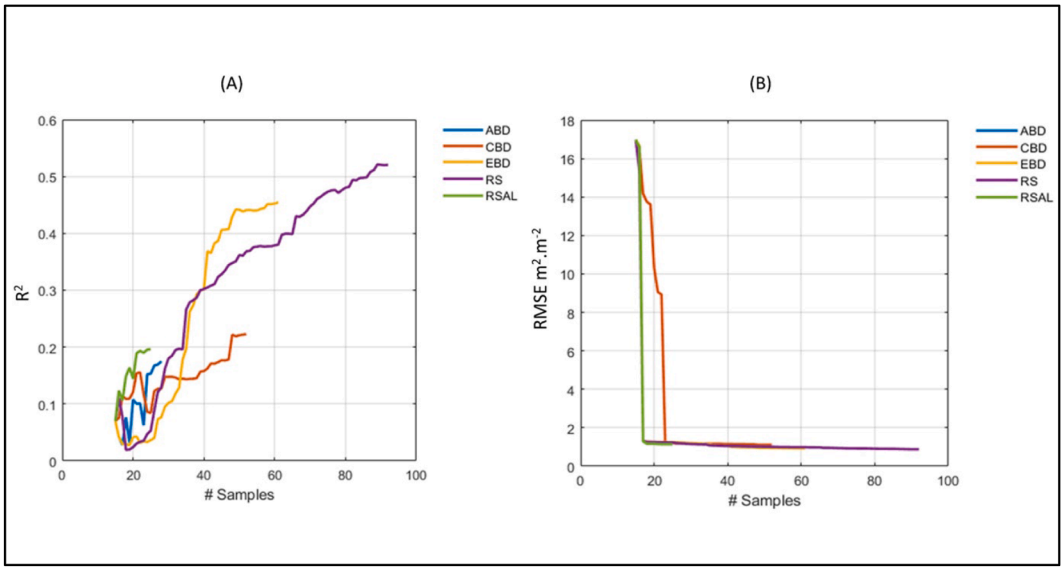

Fig. 6. Graphical representation of the R^2 and RMSE for LAI retrieval performance by KRR when trained with optimised training samples selected from a database of 1500 simulation samples by each of the five active learning (AL) algorithms.

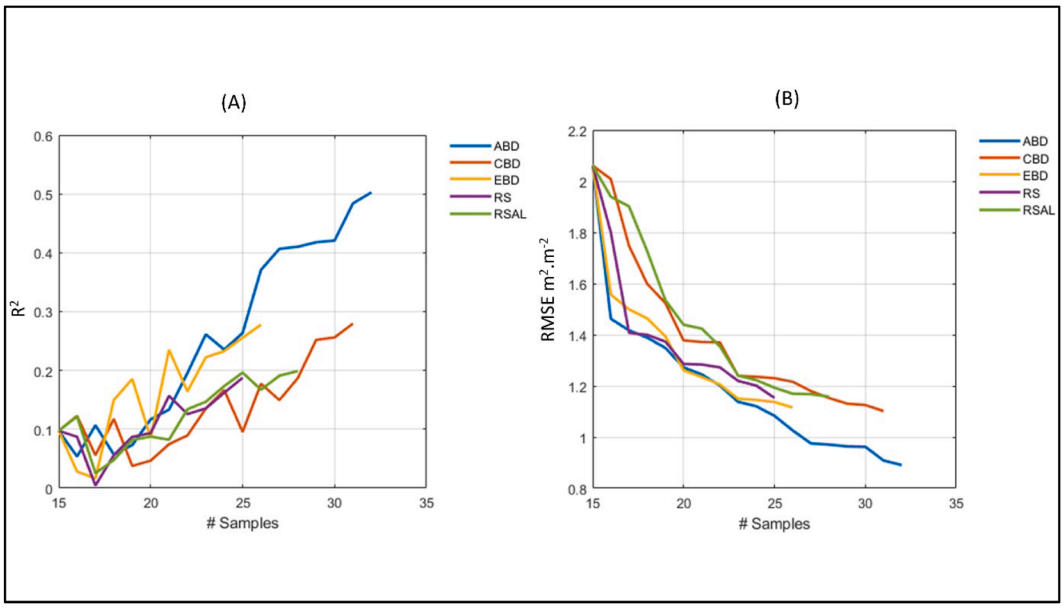

Fig. 7. Graphical representation of the R^2 and RMSE for LAI retrieval performance by the boosted regression trees (BRT) when trained with optimised training samples selected from a database of 1500 simulation samples by each of the five active learning (AL) algorithms.

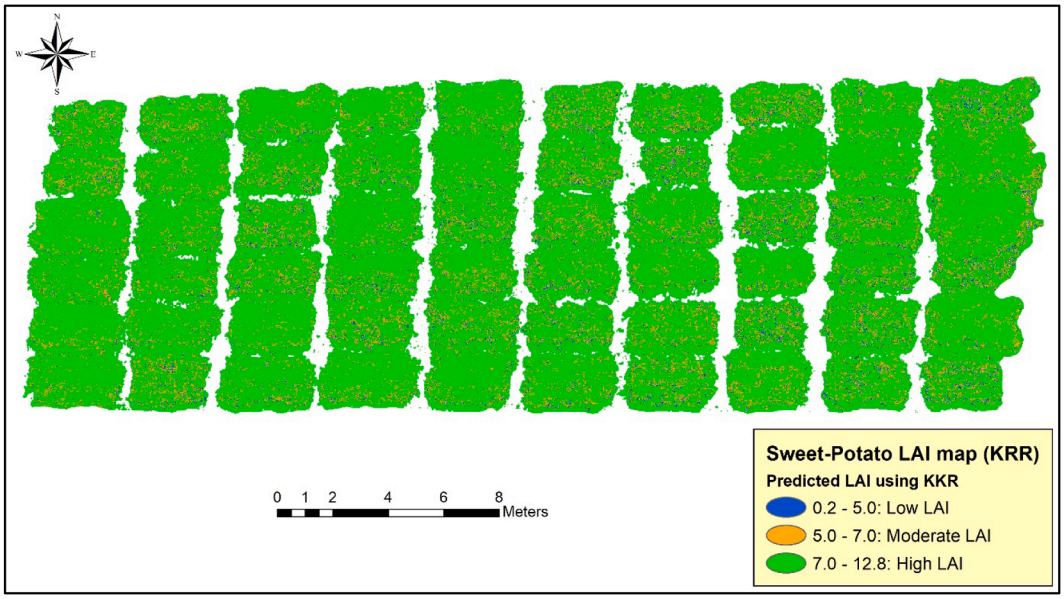

Fig. 8. Spatial prediction of LAI during matured stage (in 120 days) over multiple sweet potato varieties using the best-performing Kernel ridge regression (KRR) model integrated with the Random sampling (RS) active learning (AL) method applied to drone imagery.

range is associated with the predicted LAI mean of 7.14 $m^2.m^{-2}$ and standard deviation of 0.91 $m^2.m^{-2}$. Most LAI pixel values i.e. about 81 %, were predicted to represent high sweet potato LAI in the range 7.0 – 12.8 $m^2.m^{-2}$ which signals a largely healthy sweet potato canopy (Fig. 8). Furthermore, 15 % of the LAI pixel values were predicted to represent moderate sweet potato LAI in the range 5.0 – 7.0 $m^2.m^{-2}$ whereas about 4 % of LAI pixel values were predicted to represent low sweet potato LAI ranging from the 0.2 – 5.0 $m^2.m^{-2}$ (Fig. 8). The above distribution of LAI pixel values is also seen in Fig. 9. The distribution of predicted LAI pixel values appears approximately Gaussian with a clear depiction of the dominant LAI range (Fig. 9).

Overall, the predicted full LAI range (0.20 – 12.8 $m^2.m^{-2}$) and predicted mean (7.14 $m^2.m^{-2}$) approximated the actual field measured LAI range or minimum and maximum values of 4.31 – 9.60 $m^2.m^{-2}$ and actual field LAI mean of 7.22 $m^2.m^{-2}$ with moderately low differences (Fig. 8, Fig. 9, Table 2).

4.5. BRT spatial prediction map of sweet potato LAI

Fig. 10 shows the spatial prediction of sweet potato LAI, using the best-performing boosted regression trees (BRT) model integrated with the Angle-based diversity (ABD) active learning (AL) method i.e. R^2 of 0.50, RMSE of 0.89 $m^2.m^{-2}$ and RRMSE of 12.23 % (Table 6). In particular, the BRT model gave a spatial prediction of LAI values in the range of 2.65 – 9.32 $m^2.m^{-2}$ across the 20 sweet potato varieties at matured growth stage (Fig. 10). This range is associated with the predicted LAI mean of 7.20 $m^2.m^{-2}$ and standard deviation of 0.84 $m^2.m^{-2}$. This result closely resembled the field LAI measured mean and range of

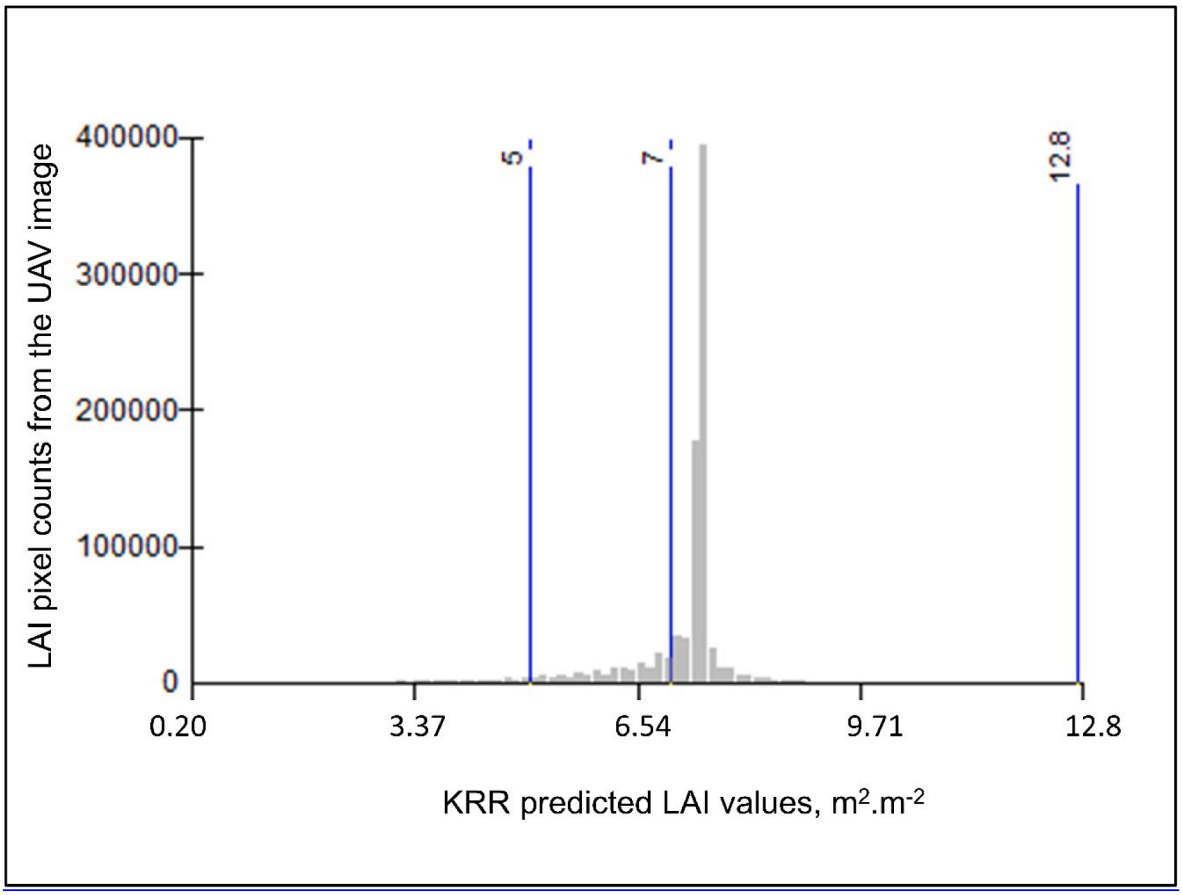

Fig. 9. Distribution of predicted LAI values from UAV imagery using the best-performing Kernel ridge regression (KRR) model integrated with the Random sampling (RS) active learning (AL) technique.

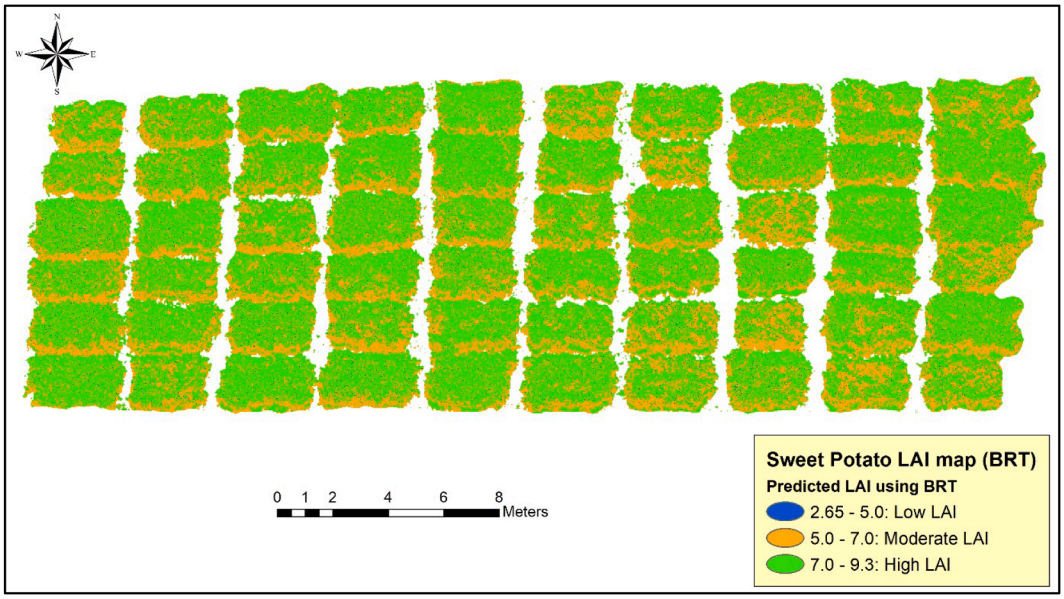

Fig. 10. Spatial prediction of LAI during matured stage (in 120 days) over multiple sweet potato varieties using the best-performing Boosted regression trees (BRT) model integrated with the Angle-based diversity (ABD) active learning (AL) method applied to drone imagery.

7.22 $m^2.m^{-2}$ and 4.31 – 9.60 $m^2.m^{-2}$ (Table 2) and this could be an indication of a more realistic LAI map shown in Fig. 10 compared to Fig. 8.

Majority of the LAI pixel values i.e. ~65 %, were predicted to represent high sweet potato LAI in the range 7 – 9.3 $m^2.m^{-2}$ which signals mainly a healthy canopy across the multiple sweet potato varieties (Fig. 10). In addition, 33 % of the LAI pixel values were predicted to represent moderate sweet potato LAI in the range 5.0 – 7.0 $m^2.m^{-2}$

whereas about 2 % of LAI pixel values were predicted to represent low sweet potato LAI ranging from the 2.65 – 5.0 $m^2.m^{-2}$ (Fig. 10). The distribution of LAI pixel values predicted by the BRT model is presented in Fig. 11 which suggests a bimodal distribution. The BRT model was able to capture well the variability of LAI, particularly between moderate and high sweet potato LAI across the canopy. Low LAI pixels are hardly visible on the predicted map.

Although, the comparison of model performance metrics in Table 4 and Table 6 showed KRR was slightly superior to BRT. However, when it comes to spatial prediction of LAI, the BRT model gave a more realistic spatial representation of predicted LAI across the multiple sweet potato varieties.

5. Discussion

This study successfully tested the integration of active learning (AL) optimization and non-parametric regression methods (NPRMs) with PROSAIL radiative transfer models (RTMs) applied to a 2-cm resolution UAV imagery, to retrieve leaf area index (LAI) at plot-scale over 20 sweet potato varieties during a matured growth stage. The results demonstrated that the NPRMs, particularly the kernel ridge regression (KRR) and boosted regression trees (BRT) have the potential to achieve accurate retrievals of sweet potato LAI when trained using a small pool of synthetic samples i.e. 1500 of RTM reflectance data over the heterogeneous canopy comprising 20 sweet potato varieties. These findings suggest that, working with a large database of synthetic training samples (e.g. of 9000 tested in this study), potentially creates largely-redundant samples for modelling the LAI of multiple sweet potato varieties in the current study. A similar observation was made in other studies (e.g. [35, 37,57,59]). For example, Li et al. [35] tested a number of PROSAIL-based reflectance training samples ranging from 1000 up to 50,000 and found that only 10,000 samples yielded the desirable estimation accuracy of LAI potato varieties. Furthermore, Ma et al. [37] added that the adequate training sample size required to give accurate retrieval of vegetation traits such LAI and canopy chlorophyll content (CCC), would vary based on growth stage and the crop variety of interest linked to their field-based growth parameter(s) required to parameterise the RTMs.

The LAI retrieval performance by KRR and BRT integrated with AL methods over 20 sweet potato varieties at peak growth stage was virtually similar. The best performing model by KRR integrated with the random sampling (RS) AL technique achieved a moderate R^2 value of 0.52 (Table 4) suggesting that, this model could explain or capture about 52 % of LAI variability across the heterogeneous sweet potato canopy at a plot-scale. Additionally, this model had the lowest prediction error based on the achieved RMSE and RRMSE values of 0.88 $m^2.m^{-2}$ and 12.23 % respectively (Table 4). On the other hand, the best performing model by BRT integrated with the angle-based diversity (ABD) AL technique achieved a moderate R^2 value of 0.50 (Table 6) which indicated that this model could explain about 50 % of LAI variability. This model had a relatively low prediction error based on the achieved RMSE and RRMSE value of 0.89 $m^2.m^{-2}$ and 12.34 % (Table 6). The LAI prediction errors of both the KRR and BRT models, fell within the acceptable range which is representative of a better LAI prediction model i.e. 0.5 $m^2.m^{-2} \leq$ RMSE $<$ 1.0 $m^2.m^{-2}$ according to Richter et al. [47]. Although, the relative performance between the KRR and BRT models showed very small differences, the spatial prediction results suggest that BRT gave a better prediction map of LAI than KRR (Fig. 10). The BRT model was able to capture well the variability of LAI, particularly between moderate and high sweet potato LAI across the canopy at a matured growth stage.

While both KRR and BRT can handle nonlinear relationships between variables, the KRR appeared to give better regression model accuracy (Table 4) than BRT (Table 6). On the other hand, BRT gave better and more realistic spatial prediction of sweet potato LAI (Fig. 8) compared to KRR (Fig. 10). In particular, BRT had predictions of the mean LAI as well as the minimum and maximum LAI range that were closest to the LAI field observations. This comparative performance between KRR and BRT could be attributed to large number (1500 and 9000) of simulations samples used for regression. For example, the algorithmic strengths of BRT is its ability to deal with large complex and noisy datasets to achieve optimal performance [20], whereas in

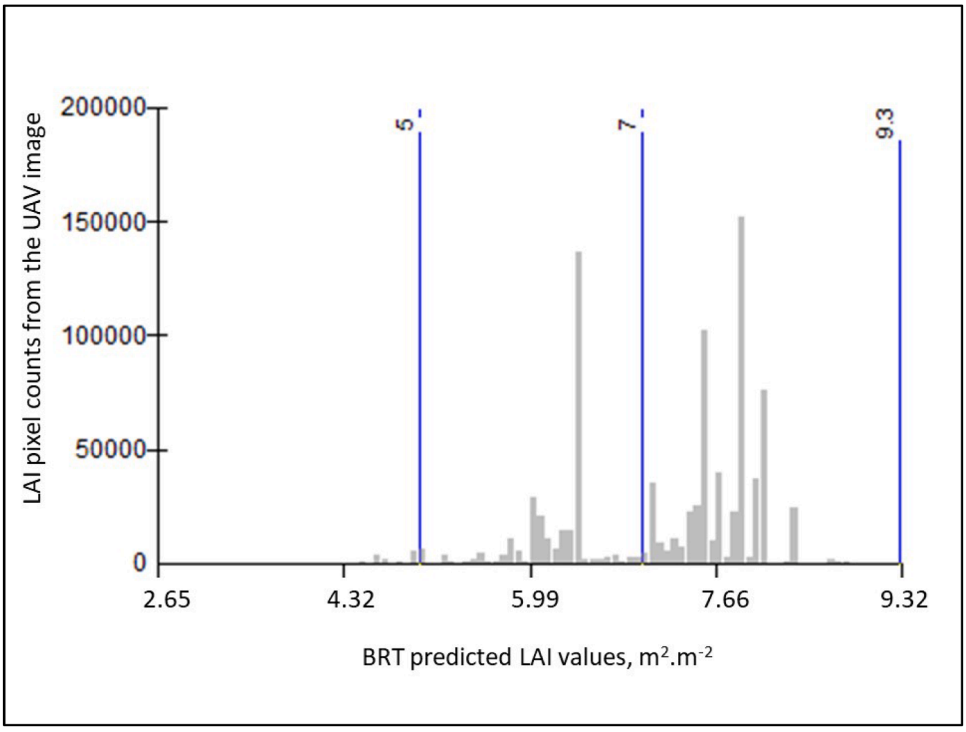

Fig. 11. Distribution of predicted LAI values from UAV imagery using the best-performing Boosted regression trees (BRT) model integrated with the Angle-based diversity (ABD) active learning (AL) technique.

instances where there is small-to-medium sized datasets with smooth relationships, the KRR may give better predictive accuracies. However, KRR's prediction accuracy can deteriorate with large or noisy datasets or large datasets with redundant samples e.g. RTM data [22].

There are very few studies based on empirical or statistical modelling approaches using LAI as a proxy for understanding the morphological and/or physiological characteristics of the sweet potato varieties at different growth stages. For example, Delazari et al. [13] used field and lab-based measurements of two sweet potato cultivars across four phonological stages based on plant growth periods to model the functional relationship between sweet potato morpho-physiological parameters (such as, LAI, leaf chlorophyll index and leaf temperate) and growth. In particular, the LAI was found to be key parameter in explaining how the variation in soil moisture and water depth influence the growth performance of the sweet potato cultivars. In another similar study, Laurie et al. [33] evaluated the growth performance of four orange-fleshed sweet potato varieties under three different irrigation treatments. Results showed that variation in levels of irrigation had a direct impact on sweet potato canopy development, and therefore, this was quantitatively expressed in the LAI values. Furthermore, there is a notable steady increase in studies that use empirical or statistical approaches to model sweet potato yield estimates and also to explore how external variables affect the shape and size properties of sweet potatoes, as these are important in determining grade and monetary value ([1,3], Liu et al. [36], Tedesco et al. [54], Tedesco et al. [55]).

In this study, an older version of the PROSPECT RTM model i.e. PROSPECT-4 [24] coupled with the SAIL model were used to model the spectral variation of sweet potato canopy reflectance, and also to generate simulated reflectance of varying sample sizes used for training the regression models. The parameterization of the aforementioned RTM models was largely based on field acquired input data, and less on assumptions and borrowed-values based on literature. This approach has been shown in other studies to improve model accuracy, including the use of more recent versions of the PROSPECT model such as PROSPECT 5 and PROSPECT-D, and incorporating lab-based derived measurements such as carotenoids and anthocyanin into the parameterization of PROSAIL [35,50]. The use of more recent versions of the PROSPECT model in the current study, which requires more input parameters such as carotenoids and anthocyaninin, are worth testing to improve the obtained results. Given the excessive costs associated with field data collection in order to obtain extensive training sets, RTMs have minimum reliance on in-situ data (Goel [21]) and are known to be robust and transferrable. In particular, RTMs use the physical laws to accurately describe the spectral variation of canopy reflectance as a function of viewing and illumination geometry, canopy, including leaf and soil background characteristics [11]. Therefore, the AL-RTM hybrid retrieval approach presented in this study could potentially be transferred to other sites with virtually similar environmental setting and target variables of interest (i.e. similar sweet potato varieties at matured growth stage). Relative to RTMs, empirical approaches on the other hand, are site specific and require abundant field (training) samples. Thus, the transferability of empirical models to other locations remains a challenge [30].

Based on reviewed literature thus far, the current study is the first to explore the integrated use of RTMs with machine learning and AL algorithms on UAV imagery for the estimation of LAI over a heterogeneous sweet potato canopy i.e. comprising 20 varieties during a single growth stage. Our findings have significant implications for: (i) sweet potato breeding programmes critical for developing new cultivars in South Africa and (ii) small to large scale farmers in obtaining accurate maps of sweet potato biophysical variables, which are essential for assessing and monitoring crop growth at different stages.

6. Conclusion

This study found that, using lesser PROSAIL simulation samples showed a better LAI retrieval performance in-terms of explained variability (R^2) and model prediction error (RMSE and RRMSE) over 20 sweet potato varieties and their replicates. Although, the retrieval performance between the kernel ridge regression (KRR) and boosted regression trees (BRT) models showed very small differences, the spatial prediction results suggest that BRT gave a better prediction map of LAI than KRR. The BRT model was able to capture well the variability of LAI, particularly between moderate and high sweet potato LAI across the canopy at a matured growth stage. Overall, these findings suggests that, the integration of BRT with the angle-based diversity (ABD) AL technique based on UAV multispectral imagery may have the potential of generating sweet potato LAI maps which are critical for phenotypic assessment of multiple sweet potato varieties at different growth stages.

The methodological approach presented in this study can be applied to other crops (such as maize, potato, wheat, beetroot and sunflower) to retrieve their biophysical variables such as LAI and Chlorophyll at different growth stages using UAV or satellite imagery. However, the parameterisation of the PROSAIL RTMs and performance of the regression models may differ across different growth stages. Based on the universal notion that RTM models are transferrable between sites, the best performing model for LAI retrieval reported in this study can potentially be transferred to other sweet potato fields for (i) identification and monitoring of crop growth and health status, (ii) identifying variations in-terms of genotypes and management practices (irrigation and fertilization), and (iii) genotype selection and adaption assessment in breeding programmes.

Funding

This research was funded the National Research Foundation (NRF) of South Africa, grant number CSRP23030580913.

Data availability

We understand that the publication of the data is becoming a good practice in research. However, we plan to share all our data in future, but at this stage we are still going to further analyse it, looking at both empirical and the inversion of the physically-based models.

CRediT authorship contribution statement

Philemon Tsele: Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Abel Ramoelo:** Writing – review & editing, Conceptualization. **Lucy Moleleki:** Writing – review & editing, Conceptualization. **Sunette Laurie:** Writing – review & editing, Resources, Data curation, Conceptualization. **Whelma Mphela:** Writing – review & editing, Resources, Data curation. **Natasha Tshuma:** Writing – review & editing, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We sincerely thank (i) Innovation Africa at University Pretoria's experimental farm and (ii) Ms Mmapaseka Malebana and team at Vegetable, Industrial and Medicinal Plants campus of the Agricultural Research Council in South Africa. Special thanks to Jochem Verrelst for providing us with the ARTMO license.

References

[1] Zakaria Alam, Sarker Umakanta, A.K.T.E.R. Sanjida, K.H.A.N. MD, Roychowdhury Rajib, Alarifi Saud, Evaluation of 17 sweet potato (Ipomoea batatas L.) genotypes across five environments for high yield and stability, Turkish J. Agricult. Forestry 48 (5) (2024) 703–719.

[2] Leo. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32.

[3] Mariella Carbajal-Carrasco, Daniela Jones, Cranos Williams, Natalie Nelson, In-season sweetpotato yield forecasting using multitemporal remote sensing environmental observations and machine learning, Authorea Preprints (2024).

[4] Zoran G Cerovic, Guillaume Masdoumier, Naïma Ben Ghozlen, Gwendal Latouche, A new optical leaf-clip meter for simultaneous non-destructive assessment of leaf chlorophyll and epidermal flavonoids, Physiol. Plant 146 (3) (2012) 251–260.

[5] Tianfeng Chai, Roland R Draxler, Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature, Geosci. Model. Dev. 7 (3) (2014) 1247–1250.

[6] Erekle Chakhvashvili, Bastian Siegmann, Onno Muller, Jochem Verrelst, Juliane Bendig, Thorsten Kraska, Uwe Rascher, Retrieval of crop variables from proximal multispectral UAV image data using PROSAIL in maize canopy, Remote Sens. (Basel) 14 (5) (2022) 1247.

[7] Qiaomin Chen, Bangyou Zheng, Karine Chenu, Pengcheng Hu, Scott C Chapman, Unsupervised plot-scale LAI phenotyping via UAV-based imaging, modelling, and machine learning, Plant Phenomics. (2022).

[8] Melba M Crawford, Devis Tuia, Hsiuhan Lexie Yang, Active learning: any value for classification of remotely sensed data? Proceed. IEEE 101 (3) (2013) 593–608.

[9] Louis Daniels, Eline Eeckhout, Jana Wieme, Yves Dejaegher, Kris Audenaert, Wouter H Maes, Identifying the optimal radiometric calibration method for UAV-based multispectral imaging, Remote Sens. (Basel) 15 (11) (2023) 2909.

[10] Danner, M., M. Locherer, T. Hank, K. Richter, and EnMAP Consortium. 2015. "Measuring leaf area index (LAI) with the LI-cor LAI 2200C or LAI-2200 (+ 2200Clear Kit)–theory, measurement, problems, interpretation".

[11] Roshanak Darvishzadeh, Clement Atzberger, Andrew Skidmore, Martin Schlerf, Mapping grassland leaf area index with airborne hyperspectral imagery: a comparison study of statistical approaches and inversion of radiative transfer models, ISPRS J. Photogramm. Remote Sens. 66 (6) (2011) 894–906.

[12] Glenn. De'Ath, Boosted trees for ecological modeling and prediction, Ecology. 88 (1) (2007) 243–251.

[13] Fábio T Delazari, Igor R Assis, Diego FV Cabrera, Mariane G Ferreira, Luiz E Dias, Angelica Rueda, Jose C Zanuncio, Derly JH Silva, Morpho-physiological characteristics by sweet potato cultivars as function of irrigation depth, Anais da Academia Brasileira de Ciências 90 (04) (2018) 3541–3549.

[14] Begüm Demir, Claudio Persello, Lorenzo Bruzzone, Batch-mode active-learning methods for the interactive classification of remote sensing images, IEEE Transac. Geosci. Remote Sens. 49 (3) (2010) 1014–1031.

[15] Yadong Dong, Ziti Jiao, Lei Cui, Hu Zhang, Xiaoning Zhang, Siyang Yin, Anxin Ding, Yaxuan Chang, Rui Xie, Jing Guo, Assessment of the hotspot effect for the PROSAIL model with POLDER hotspot observations based on the hotspot-enhanced kernel-driven BRDF model, IEEE Transac. Geosci. Remote Sens. 57 (10) (2019) 8048–8064.

[16] F. Douak, N. Benoudjit, Farid Melgani, A two-stage regression approach for spectroscopic quantitative analysis, Chemometr. Intellig. Laborat. Syst. 109 (1) (2011) 34–41.

[17] Fouzi Douak, Farid Melgani, Nabil Benoudjit, Kernel ridge regression with active learning for wind speed prediction, Appl. Energy 103 (2013) 328–340.

[18] Si-Bo Duan, Zhao-Liang Li, Hua Wu, Bo-Hui Tang, Lingling Ma, Enyu Zhao, Chuanrong Li, Inversion of the PROSAIL model to estimate leaf area index of maize, potato, and sunflower fields from unmanned aerial vehicle hyperspectral data, Int. J. Appl. Earth Observ. Geoinform. 26 (2014) 12–20.

[19] J.B. Féret, G. Le Maire, S. Jay, D. Berveiller, R. Bendoula, G. Hmimina, A. Cheraiet, J.C. Oliveira, F.J. Ponzoni, T. Solanki, F. De Boissieu, Estimating leaf mass per area and equivalent water thickness based on leaf optical properties: potential and limitations of physical modeling and machine learning, Remote Sens. Environ. 231 (2019) 110959.

[20] Steven K. Friedman, Janet Franklin (With a Contribution By Jennifer A. Miller): Mapping species distributions: Spatial inference and Prediction, 2009, Cambridge University Press, 2011, p. 320. Illus, maps; Hardback, ISBN–978-0-521-87635-3 (122.00), Paperback ISBN 978-0-521-70002-3, US $53.00." In.: Springer.

[21] Goel, Narendra S. 1987. "Inversion of canopy reflectance models for estimation of vegetation parameters." In.

[22] Trevor Hastie, Robert Tibshirani, Jerome H Friedman, Jerome H Friedman, The Elements of Statistical learning: Data mining, inference, and Prediction, 2, Springer, 2009.

[23] Alexandre Bryan Heinemann, Pepijn AJ Van Oort, Diogo Simões Fernandes, Aline de Holanda Nunes Maia, Sensitivity of APSIM/ORYZA model due to estimation errors in solar radiation, Bragantia 71 (2012) 572–582.

[24] Stéphane Jacquemoud, Frédéric Baret, PROSPECT: a model of leaf optical properties spectra, Remote Sens. Environ. 34 (2) (1990) 75–91.

[25] Stéphane Jacquemoud, Wout Verhoef, Frédéric Baret, Cédric Bacour, Pablo J Zarco-Tejada, Gregory P Asner, Christophe François, Susan L Ustin, PROSPECT+ SAIL models: a review of use for vegetation characterization, Remote Sens. Environ. 113 (2009) S56–S66.

[26] Anil K Jain, Richard C Dubes, Algorithms For Clustering Data, Prentice-Hall, Inc, 1988.

[27] P.D. Jamieson, J.R. Porter, D.R. Wilson, A test of the computer simulation model ARCWHEAT1 on wheat crops grown in New Zealand, Field. Crops. Res. 27 (4) (1991) 337–350.

[28] Sylvain Jay, Frédéric Baret, Dan Dutartre, Ghislain Malatesta, Stéphanie Héno, Alexis Comar, Marie Weiss, Fabienne Maupas, Exploiting the centimeter resolution of UAV multispectral imagery to improve remote-sensing estimates of canopy structure and biochemistry in sugar beet crops, Remote Sens. Environ. 231 (2019) 110898.

[29] Inge Jonckheere, Stefan Fleck, Kris Nackaerts, Bart Muys, Pol Coppin, Marie Weiss, Frédéric Baret, Review of methods for in situ leaf area index determination: part I. Theories, sensors and hemispherical photography, Agric. For. Meteorol. 121 (1–2) (2004) 19–35.

[30] Mahlatse Kganyago, Clement Adjorlolo, Paidamwoyo Mhangara, Exploring transferable techniques to retrieve crop biophysical and biochemical variables using sentinel-2 data, Remote Sens. (Basel) 14 (16) (2022) 3968.

[31] A. Kuusk, The hot spot effect in plant canopy reflectance. Photon-vegetation interactions: Applications in Optical Remote Sensing and Plant Ecology, Springer, 1991, pp. 139–159.

[32] R.N. Laurie, S.M. Laurie, C.P. Du Plooy, J.F. Finnie, J. Van Staden, Yield of drought-stressed sweet potato in relation to canopy cover, stem length and stomatal conductance, J. Agricult. Sci. 7 (1) (2015) 201.

[33] R.N. Laurie, CP du Plooy, S.M. Laurie, Effect of moisture stress on growth and performance of orange fleshed sweetpotato varieties, in: African Crop Science Conference Proceedings 9, African Crop Science Society, Cape Town, South Africa, 2009, pp. 235–239, 28 September - 2 October 2009, ISSN 1023-070X/2009.

[34] Sunette Laurie, Frikkie Calitz, Musa Mtileni, Whelma Mphela, Sidwell Tjale, Performance of informal market sweet potato cultivars in on-farm trials in South Africa, Open. Agric. 2 (1) (2017) 431–441.

[35] Shuang Li, Yongxin Lin, Ping Zhu, Liping Jin, Chunsong Bian, Jiangang Liu, Combining UAV multispectral imaging and PROSAIL model to estimate LAI of potato at plot scale, Agriculture 14 (12) (2024) 2159.

[36] Hangjin Liu, Shelly Hunt, G. Craig Yencho, Kenneth V Pecota, Russell Mierop, Cranos M Williams, Daniela S Jones, Predicting sweetpotato traits using machine learning: impact of environmental and agronomic factors on shape and size, Comput. Electron. Agric. 225 (2024) 109215.

[37] Yuanyuan Ma, Xiaoyu Song, Jie Zhang, Di Pan, Haikuan Feng, Guijun Yang, Chunxia Qiu, Heguang Sun, Chunkai Zheng, Pingping Li, Inversion of biophysical parameters of potato based on an active learning pool-based sampling strategy, Int. J. Remote Sens. 46 (1) (2025) 320–339.

[38] David JC. MacKay, Information-based objective functions for active data selection, Neural Comput. 4 (4) (1992) 590–604.

[39] Madodomzi Mafanya, Philemon Tsele, Joel O Botai, Phetole Manyama, George J Chirima, Thabang Monate, Radiometric calibration framework for ultra-high-resolution UAV-derived orthomosaics for large-scale mapping of invasive alien plants in semi-arid woodlands: harrisia pomanensis as a case study, Int. J. Remote Sens. 39 (15–16) (2018) 5119–5140.

[40] Maria Jose Marques, Ana Alvarez, Pilar Carral, Blanca Sastre, Ramón Bienes, The use of remote sensing to detect the consequences of erosion in gypsiferous soils, Int. Soil Water Conserv. Res. 8 (4) (2020) 383–392.

[41] Nozipho M Motsa, Albert T Modi, Tafadzwanashe Mabhaudhi, Sweet potato (Ipomoea batatas L.) as a drought tolerant and food security crop, S. Afr. J. Sci. 111 (11–12) (2015) 1–8.

[42] Joaquín Muñoz-Sabater, Emanuel Dutra, Anna Agustí-Panareda, Clément Albergel, Gabriele Arduini, Gianpaolo Balsamo, Souhail Boussetta, Margarita Choulga, Shaun Harrigan, Hans Hersbach, ERA5-Land: a state-of-the-art global reanalysis dataset for land applications, Earth. Syst. Sci. Data 13 (9) (2021) 4349–4383.

[43] Edoardo Pasolli, Farid Melgani, Naif Alajlan, Yakoub Bazi, Active learning methods for biophysical parameter estimation, IEEE Transac. Geosci. Remote Sens. 50 (10) (2012) 4071–4084.

[44] Swarnajyoti Patra, Lorenzo Bruzzone, A cluster-assumption based batch mode active learning technique, Pattern. Recognit. Lett. 33 (9) (2012) 1042–1048.

[45] Peterson, P., C.C. Funk, G.J. Husak, D.H. Pedreros, M. Landsfeld, J.P. Verdin, and S. Shukla. 2013. The climate hazards group infrared precipitation (chirp) with stations (chirps): development and validation. Paper presented at the AGU Fall meeting abstracts.

[46] David A Ramírez, Wolfgang Grüneberg, Maria I Andrade, Bert De Boeck, Hildo Loayza, Godwill S Makunde, Johan Ninanya, Javier Rinza, Simon Heck, Hugo Campos, Phenotyping of productivity and resilience in sweetpotato under water stress through UAV-based multispectral and thermal imagery in Mozambique, J. Agron. Crop. Sci. 209 (1) (2023) 41–55.

[47] Katja Richter, Clement Atzberger, Tobias B Hank, Wolfram Mauser, Derivation of biophysical variables from Earth observation data: validation and statistical measures, J. Appl. Remote Sens. 6 (1) (2012) 063557. -.

[48] Saunders, Craig, Alexander Gammerman, and Volodya Vovk. 1998. "Ridge regression learning algorithm in dual variables".

[49] Glen Schloderer, Matthew Bingham, Joseph L Awange, Kevin M Fleming, Application of GNSS-RTK derived topographical maps for rapid environmental monitoring: a case study of Jack Finnery Lake (Perth, Australia), Environ. Monit. Assess. 180 (1) (2011) 147–161.

[50] Bongokuhle Sibiya, Moses Cho, Onisimo Mutanga, John Odindi, Cecilia Masemola, Wessel Bonnet, Determining the value of different wavelength ranges of non-imaging hyperspectral reflectance to estimate carotenoid content using the PROSPECT-5 model, Int. J. Remote Sens. 46 (4) (2025) 1696–1719.

[51] Ronald D. Snee, Validation of regression models: methods and examples, Technometrics. 19 (4) (1977) 415–428.

[52] Johan AK Suykens, Joos Vandewalle, Least squares support vector machine classifiers, Neural Process. Lett. 9 (1999) 293–300.

[53] Danilo Tedesco, Bruno Rafael de Almeida Moreira, Marcelo Rodrigues Barbosa Júnior, Murilo Maeda, Rouverson Pereira da Silva, Sustainable management of

sweet potatoes: a review on practices, strategies, and opportunities in nutrition-sensitive agriculture, energy security, and quality of life, Agric. Syst. 210 (2023) 103693.
[54] Danilo Tedesco, Bruno Rafael de Almeida Moreira, Marcelo Rodrigues Barbosa Júnior, João Paulo Papa, Rouverson Pereira da Silva, Predicting on multi-target regression for the yield of sweet potato by the market class of its roots upon vegetation indices, Comput. Electron. Agric. 191 (2021) 106544.
[55] Danilo Tedesco, Maílson Freire de Oliveira, Adão Felipe dos Santos, Edgard Henrique Costa Silva, Glauco de Souza Rolim, Rouverson Pereira da Silva, Use of remote sensing to characterize the phenological development and to predict sweet potato yield in two growing seasons, in: Europ. J. Agron., 129, 2021 126337.
[56] Mphethe I Tongwane, Michael J Savage, Mitsuru Tsubo, Relationship between global and diffuse irradiance and their variability in South Africa, Theor. Appl. Climatol. 137 (2019) 1027–1040.
[57] Philemon Tsele, Abel Ramoelo, Integrating active learning and regression methods for estimation of grass lai over a mountainous region using Sentinel-2 satellite data, Paper presented at, in: the IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium, 2024.
[58] Jochem Verrelst, Gustau Camps-Valls, Jordi Muñoz-Marí, Juan Pablo Rivera, Frank Veroustraete, Jan GPW Clevers, José Moreno, Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties–A review, ISPRS J. Photogramm. Remote Sens. 108 (2015) 273–290.
[59] Jochem Verrelst, Sara Dethier, Juan Pablo Rivera, Jordi Munoz-Mari, Gustau Camps-Valls, Jose Moreno, Active learning methods for efficient hybrid biophysical variable retrieval, IEEE Geosci. Remote Sens. Lett. 13 (7) (2016) 1012–1016.
[60] Jochem Verrelst, Zbyněk Malenovský, Christiaan Van der Tol, Gustau Camps-Valls, Jean-Philippe Gastellu-Etchegorry, Philip Lewis, Peter North, Jose Moreno, Quantifying vegetation biophysical variables from imaging spectroscopy data: a review on retrieval methods, Surv. Geophys. 40 (2019) 589–629.
[61] Hang Yin, Weili Huang, Fei Li, Haibo Yang, Yuan Li, Yuncai Hu, Kang Yu, Multi-temporal UAV imaging-based mapping of chlorophyll content in potato crop, PFG–J. Photogramm., Remote Sens. Geoinform. Sci. 91 (2) (2023) 91–106.