Above-ground biomass estimation and yield prediction in potato by using UAV-based RGB and hyperspectral imaging

Bo Li[a,b,c], Xiangming Xu[b], Li Zhang[b], Jiwan Han[c], Chunsong Bian[a], Guangcun Li[a], Jiangang Liu[a,*], Liping Jin[a,*]

[a] Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences (CAAS), 12 Zhongguancun, Haidian, Beijing, 100081, China

[b] NIAB EMR, New road, East malling, Kent, ME19 6BJ, UK

[c] Shanxi Agricultural University, Taigu, Shanxi, 030801, China

[*] Corresponding author. E-mail address: liujiangang@caas.cn; jinliping@caas.cn

Abstract:

Rapid and accurate biomass and yield estimation facilitates efficient plant phenotyping and site-specific crop management. A low altitude unmanned aerial vehicle (UAV) was used to acquire RGB and hyperspectral imaging data for a potato crop canopy at two growth stages to estimate the above-ground biomass and predict crop yield. Field experiments included six cultivars and multiple treatments of nitrogen, potassium, and mixed compound fertilisers. Crop height was estimated using the difference between digital surface model and digital elevation models derived from RGB imagery. Combining with two narrow-band vegetation indices selected by the RReliefF feature selection algorithm. Random Forest regression models demonstrated high prediction accuracy for both fresh and dry above-ground biomass, with a coefficient of determination (r^2) > 0.90. Crop yield was predicted using four narrow-band vegetation indices and crop height (r^2 = 0.63) with imagery data obtained 90 days after planting. A Partial Least Squares regression model based on the full wavelength spectra demonstrated improved yield prediction (r^2 = 0.81). This study demonstrated the merits of UAV-based RGB and hyperspectral imaging for estimating the above-ground biomass and yield of potato crops, which can be used to assist in site-specific crop management.

Key words: unmanned aerial vehicle; hyperspectral imaging; potato; above-ground biomass; yield prediction

1. **Introduction**

Potato (*Solanum tuberosum* L.) is the fourth most important staple food in the world. Consequently, improving potato production without negative environmental consequences is important for ensuring global food security. Above-ground biomass (AGB) is closely related to crop nutrition status and yield; hence, it can be used as an indicator of crop growth status. Understanding the spatio-temporal dynamics of AGB and its relationship to yield is essential for developing and implementing site-specific crop husbandry measures. AGB is commonly measured using manual sampling, which is extremely time-consuming (Freeman et al., 2007), while yield prediction is largely dependent on subjective, often inaccurate, and labour-intensive ground-based visits (Reynolds et al., 2000).

Remote sensing is an efficient technique for measuring growing season crop canopies and to provide information on the spatial variability of crop AGB and yield. RGB imaging is a low-cost solution that can be used for AGB estimation. For example, Bendig et al. (2014) calculated crop height using a digital surface model (DSM) derived from unmanned aerial vehicle (UAV) based RGB imaging as an indicator of AGB; however, model accuracy was cultivar dependent. In addition to crop height, canopy cover and volume were found to be good predictors of onion dry bulb biomass (Ballesteros et al., 2018). For example, a vegetation index (VI) weighted canopy volume model incorporating canopy area, height, and VIs derived from RGB imaging produced an accurate prediction of soybean biomass for different genotypes (Maimaitijiang et al., 2019). With the use of spectral imaging sensors in agriculture, VIs are commonly used to estimate AGB and predict yields for wheat (Raun et al., 2001; Yue et al., 2017), barley (Hansen et al., 2002; Tilly et al., 2015), maize (Gitelson et al., 2003; Shanahan et al., 2001), rice (Swain et al., 2010), and cotton (Bai et al., 2007; Zhao et al., 2007).

Single broad-band VIs, such as the normalised difference vegetation index (NDVI), employ limited spectral information (Mutanga and Skidmore, 2004); thus, multiple VIs are commonly combined as predictor variables. For example, a random forest (RF) model based on multiple broad-band VIs was developed to estimate wheat biomass by Wang et al. (2016) and was found to perform better than both artificial neural network (ANN) and support vector regression (SVR) models. Similarly, an RF model derived using VIs and crop height-related metrics from a crop surface model was able to predict maize biomass with a slightly higher accuracy than ANN and SVR models (Han et al., 2019). Narrow-band VIs (Haboudane et al., 2002) have been developed to utilise hyperspectral sensors and powerful data mining techniques. A partial least squares (PLS) regression model based on all pairwise two-band NDVI combinations predicted wheat AGB satisfactorily (Hansen and Schjoerring, 2003). A range of VIs was selected using a support vector machine (SVM) and the weighted difference vegetation index was found to have the best predictive power for grassland AGB (Clevers et al., 2007). Due to the lack of three-dimensional (3D) canopy structure information, there are difficulties in using spectral imaging exclusively to estimate the plant biomass of various heights and densities (Greaves et al., 2015). For example, a fused multivariate model with plant height and narrow-band VIs was introduced to predict barley biomass, and showed better performance than using VIs only (Tilly et al., 2015). Currently, there is limited research regarding the use of remote sensing to estimate the AGB of potato. The cumulative ratio of the radiance of the near-infrared and red bands was related to potato crop dry biomass; however, such a relationship was dependent on crop nitrogen (N) status (Millard et al., 1990).

Remote sensing methods for crop yield prediction currently rely on broad-band VIs such as the NDVI (Huang et al., 2013; Prasad et al., 2006; Raun et al., 2001; Vergara-Díaz et al., 2016). While the NDVI is related to yield, it can be influenced by other factors, including the soil background and light conditions (Thenkabail et al., 2016). Consequently, other broad-band VIs have also been used as indicators for crop yield. For example, the area of the red edge peak was correlated with wheat grain yield by Cao et al. (2015), while a simple ratio had a higher correlation with wheat yield compared to NDVI and the photochemical reflectance index (Aparicio et al., 2000). Furthermore, the green normalised difference vegetation index was highly correlated with corn grain yield (Shanahan et al., 2001). Using hyperspectral sensors, there are more narrow-band VIs available for yield prediction. Both stepwise multiple linear regression (MLR) and ANN models based on narrow-band VIs predicted corn yield well (Uno et al., 2005). However, narrow-band VIs lose a large amount of spectral information, which may explain why yield predictive models based on these VIs are often cultivar specific (Montesinos-López et al., 2017). A chemometric analysis using all bands as predictor variables improved the prediction accuracy over using VIs alone for wheat yield prediction (Montesinos-López et al., 2017). Furthermore, improved predictive performance was achieved for citrus yield using a PLS model with all bands compared to MLR models with narrow-band VIs (Ye et al., 2007). For potato yield prediction, a soil adjusted vegetation index derived from satellite imagery was found to correlate with potato yield (Al-Gaadi et al., 2016). The red-edge chlorophyll index 1 (CI1) predicted total potato yield as early as 55 days after planting (DAP) with a reasonable accuracy (Morier et al., 2015). However, there is limited research using multiple VIs or the full spectra from UAV-based hyperspectral imaging to predict potato yield.

Compared with ground-based and satellite-based remote sensing techniques, UAV-based imaging can achieve satisfactory temporal, spatial, and spectral resolution (Sankaran et al., 2015). This study applies UAV-based RGB and hyperspectral imaging to: (1) compare estimations of crop height using the DSM-based method and the full spectra PLS regression model; (2) predict AGB using the RF model with VIs and crop height, and compare the performance of the RF model with the full spectra PLS regression model; and (3) predict yield using the RF model with crop height and VIs, and compare the performance of the RF model with the full spectra PLS regression model.

2. Materials and methods

2.1. Study Area

Three experiments were conducted at the Chinese Academy of Agricultural Sciences research station located in Zhangjiakou, Hebei, China (41º 28 ′28.82 ″N, 115º 03 ′43.91 ″E, and elevation 1390 m). Experiments varied input levels of N, K, and mixed organic-inorganic compound fertilisers to generate different levels of AGB and yield (Fig. 1). Seed potatoes were sown on the 6th May 2018 and harvested on the 10th September 2018. Experiment 1 was comprised of five blocks, each with a different N input level (0, 100, 200, 300, and 400 kg ha^{-1}). Within each block, there were twelve plots, each sown with one of four cultivars including *Favorita*, *Zhongshu10* (Z10), *Zhongshu18* (Z18), and *Zhongshu19* (Z19). Experiment 2 contained three blocks, with 12 plots per block. Each block had different K input levels (0, 75, 150, and 225 kg ha^{-1}) with cultivars including *Zhongshu5* (Z5), Z18, and *Shepody*. Experiment 3 contained three blocks, with 16 plots per block. Each block was comprised of a combination of eight different mixed compound fertilisers (see A1 for the details of the mixed compound fertilisers) and two cultivars (Z5 and Z18).

The plot size in Experiments 1 and 2 was 8 x 5.3 m, containing six rows with 270 evenly sown seed potatoes. In Experiment 3, plot size was 6 x 5.3 m containing six rows with 210 evenly sown seed potatoes. Of the six cultivars used, *Favorita*, Z5, and Z10 are early maturing, while Z18, Z19, and *Shepody* are late maturing. A selective herbicide (DuPont Matrix) was applied at the emergence stage to minimise the effect of weeds on image analysis.

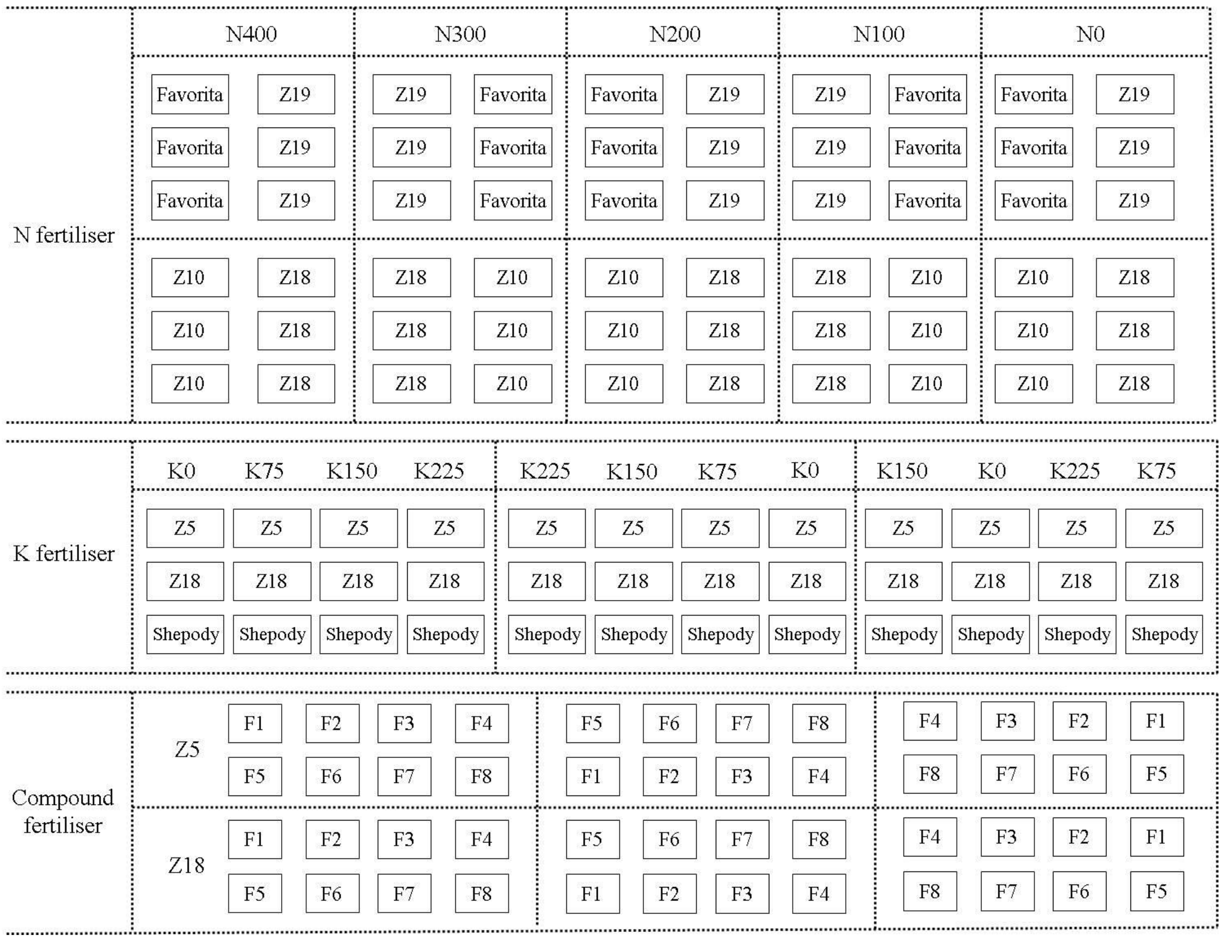


Figure 1. Layout of the experimental plots. Potato cultivars *Favorita*, *Shepody*, *Zhongshu5* (Z5), *Zhongshu10* (Z10), *Zhongshu18* (Z18), and *Zhongshu19* (Z19) were planted in three experimental fields receiving different N, K, and mixed compound fertiliser treatments. The details of the eight different mixed compound fertilisers are shown in A1.

2.2. Image acquisition and pre-processing

Both RGB and hyperspectral imaging data were obtained under clear sky conditions on the 5th July and 6th August 2018, approximately 60 and 90 DAP, respectively. The RGB images were taken by a lightweight UAV (DJI Phantom 4 Pro) equipped with a 20 mega pixel camera at a flight altitude of 30 m, equivalent to a spatial resolution of 0.5 cm/pixel. The flight survey was configured with a 60% side and 80% forward overlap. The imagery and corresponding position and orientation system (POS) data were used to generate an orthomosaic image and a DSM of the site using Pix4d software (Lausanne, Switzerland) and the structure from motion (SfM) algorithm (Colomina and Molina, 2014). The August

RGB orthomosaic image was then co-registered to the July RGB orthomosaic image based on 30 unchanged ground features, including fixed irrigation connections and physical markers, using ENVI 5.3 software (Research Systems Inc., Boulder Co., USA). Hyperspectral imaging data were captured at a flight altitude of 30 m with 60% side overlap by a DJI Matrice 600 Pro Hexacopter equipped with a Headwall Nano-Hyperspec (Headwall Photonics Inc., Bolton, MA, USA) push-broom sensor that offers 272 spectral bands and 640 spatial pixels within the visible-near-infrared range from 400–1000 nm. The spatial resolution of the hyperspectral images obtained on the two flight dates differed slightly; 2.2 cm/pixel for the first flight survey and 3.1 cm/pixel for the second. Radiometric and geometric corrections were applied to raw image strips using corresponding onboard navigation information and in-situ grey-white reflectance calibration panels for each flight to produce georeferenced reflectance images. Each calibrated image strip was then co-registered to their corresponding RGB orthomosaic imagery with at least 20 ground control points (GCPs) using the nearest neighbour resampling method (with second degree polynomial interpolation) in ENVI and Interactive Data Language (IDL). Due to the different flight directions and image spatial resolutions between the two surveys, 16 July image strips were processed to produce a mosaic image covering the field while 9 longer and slightly lower resolution image strips were used for the August mosaic. Fixing points including irrigation pipes, coloured field markers, and small but distinct green vegetation in between rows were identified from both RGB and hyperspectral image strips as GCPs. Between 17–25 GCPs evenly distributed across imagery were used for each July image with an average resampling root mean squared error (RMSE) of 2.3 cm (0.71–1.43 pixels). Two of the 16 image strips were divided into two sub-images through the wide gap between plots and rectified separately to avoid high RMSE in the crop areas of the image. In the August imagery, the potato canopy in most of the plots was closed; crop rows had merged, and some of the markers were covered by the crop canopy. Less obvious points were identified between crop gaps. Insufficient GCPs were identified in the fertiliser experiment plots to ensure an even distribution of GCPs in the image. Instead, selected small clusters of potato flowers were used as GCPs. Because the image strips are longer in August, 28–39 GCPs were used for each image. The second degree polynomial nearest neighbour resampling method was used and yielded very good rectification results with an average RMSE of 2.2 cm (0.31–0.82 pixels). The fine-tuned rectified image strips were then used to produce a hyperspectral mosaic of the field site using the ENVI mosaic tool. A seamline was designed for each image following crop gaps and 10 pixel feathering was applied to the overlapping area of neighbouring image strips. All edges of the image strips with larger RMSE were removed during mosaicing. The hyperspectral image mosaic showed strong agreement with the corresponding RGB images.

2.3. Field crop assessment

Field measurements were conducted on the same days as the UAV surveys (5th July and 6th August 2018) to provide ground truth data. Three plants were randomly selected at the centre of each plot and their heights were measured with a telescopic levelling rod. The average height of the three plants was then used to represent the canopy height of each plot. The fresh AGB of another three randomly selected plants at the centre of each plot from the N fertiliser experiment (Experiment 1) was obtained on the same day. The corresponding dry weight was obtained after the fresh samples were dried at 80 ºC for 48 h. The AGB per hectare was calculated by:

$$AGB = AGB_{ave} \times n \tag{1}$$

where AGB_{ave} is the average biomass of potato plant samples and n is the number of potato plants per hectare estimated using plot plant density. Similarly, yield data were measured by weighing the total weight of potato tubers within each plot. These conversions were necessary because the plot size in Experiment 3 differed from the other two experiments.

2.4. Image processing and data extraction

2.4.1. Extraction of spectra from hyperspectral images

To extract the spectra corresponding to the green canopy, it was necessary to generate a binary mask image by segmenting the green canopy from the soil background. The excessive green index (ExG) was a robust VI, facilitating contrast enhancement between the potato canopy and soil background (Li et al., 2019) as follows:

$$ExG_{xy} = 2R_{540} - R_{465} - R_{680} \quad (2)$$

where R_{465}, R_{540}, and R_{680} are the reflectance intensities at 465, 540, and 680 nm, in the blue, green and red regions, respectively, and x and y are the coordinates of a specific pixel. The Otsu thresholding method (Otsu, 1979) was applied to convert the ExG greyscale image to a binary image with a zero value assigned to soil background, and the spectra were extracted from non-zero pixels as a region of interest. An average spectra value was calculated for each plot.

2.4.2. Vegetation indices

VIs are mathematical transformations of the spectra at pre-defined wavelengths. With the use of hyperspectral sensors, many narrow-band VIs have been developed in recent years for estimating crop biophysical parameters (Silleos et al., 2006). Several VIs have been applied to potato crops for estimating leaf chlorophyll, leaf area index, ground cover (Domingues Franceschini et al., 2017), N content (Herrmann et al., 2010; Jain et al., 2007), and yield (Morier et al., 2015). Based on these studies (Clark et al., 2011; Yue et al., 2017), 13 VIs (Table 1) that showed good correlations with biophysical parameters, potato crop yield, and the biomass of other crops were selected for use in this study.

Table 1. Narrow-band vegetation indices (VIs) used in this study.

Vegetation index	Equation	Reference
NDVI (normalized difference vegetation index)	$NDVI=(R_{850}-R_{680})/(R_{850}+R_{680})$	Rouse et al. (1974)
MSR (modified simple ratio)	$MSR=(R_{800}-R_{670}-1)/[(R_{800}+R_{670})0.5+1]$	Chen et al. (1996)
MSAVI (modified soil adjusted vegetation index)	$MSAVI=R_{800}+0.5-[(R_{800}+0.5)2-2(R_{800}-R_{670})]^{0.5}$	Qi et al. (1994)
OSAVI (optimised soil adjusted vegetation index)	$OSAVI=(1+0.16)(R_{800}-R_{670})/(R_{800}+R_{670}+0.16)$	Rondeaux et al. (1996)
MCARI (modified chlorophyll absorption reflectance index)	$MCARI=[(R_{700}-R_{600})-0.2(R_{700}-R_{550})](R_{700}/R_{670})$	Daughtry et al. (2000)
MCARI2	$MCARI2=1.5[2.5(R_{800}-R_{670})-1.3(R_{800}-R_{550})]/[(2R_{800}+1)2-(6R_{800}-5R_{670}^{0.5})-0.5]$	Haboudane et al. (2004)

TCARI (transformed chlorophyll absorption reflectance index)	TCARI=3[(R_{700}-R_{670})-0.2(R_{700}-R_{550})(R_{700}/R_{670})]	Haboudane et al. (2002)
NDI (normalized difference index)	NDI=(R_{850}-R_{710})/(R_{850}+R_{680})	Datt et al. (1999)
CI1 (red-edge chlorophyll index 1)	CI1=R_{800}/R_{740}-1	Li et al. (2012)
CI2 (red-edge chlorophyll index 2)	CI2=R_{740}/R_{550}-1	Gitelson et al. (1996)
SIPI (structure-insensitive pigment index)	SIPI=(R_{800}-R_{445})/(R_{800}+R_{680})	Penuelas et al. (1995)
TCARI/OSAVI	TCARI/OSAVI	Haboudane et al.(2002)
MCARI/OSAVI	MCARI/OSAVI	Zarco-Tejada et al. (2004)

2.5. Data analysis

2.5.1. RReliefF *algorithm for feature selection*

Not all predictor variables are equally important to a machine learning model, and redundant variables can markedly reduce model performance (Son et al., 2015). Selection of the optimal predictor variables in this study was based on the RReliefF algorithm (Kira and Rendell, 1992), also known as the regression version of ReliefF. RReliefF introduces probabilities that can be modelled by the relative distance between the predicted values of two observations, and can calculate the quality weights of all variables as shown in Fig. 2:

RReliefF algorithm

Input: Training instance x_k with *F* variables;
m = number of samples; *k* = number of nearest neighbours
Output: *W* - Quality weight vector for all variables

Initialise N_{dC}, and all elements in N_{dA}, $N_{dC\wedge dA}$, *W* to 0;
for *i* = 1 **to** *m* **do**:
  select instance R_i randomly;
  select *k* nearest instances I_j to R_i;
  for *j* = 1 **to** *k* **do**:
    # index 0 in diff function refers to target variable
    $N_{dC} = N_{dC} + diff(0, I_j, R_i)/k;$
    for *A* = 1 **to** *F* **do**:
      $N_{dA}(A) = N_{dA}(A) + diff(A, I_j, R_i)/k;$
      $N_{dC\wedge dA}(A) = N_{dC\wedge dA}(A) + diff(0, I_j, R_i) * diff(A, I_j, R_i)/k;$
    end
  end
end
for *A* = 1 **to** *F* **do**:
  $W(A) = N_{dC\wedge dA}(A) / N_{dC} - (N_{dA}(A) - N_{dC\wedge dA}(A)) / (m - N_{dC});$

Figure 2. Explanation of the RReliefF algorithm in pseudo code.

where:

$$diff(A,\ I_j,\ R_i) = \frac{|value(A, I_j) - value(A, R_i)|}{(A_{max} - A_{min})} \tag{3}$$

$value(A,\ I_j)$ is the value of A attributes for samples I_j and R_i, and A_{max} and A_{min} are the maximum and minimum values, respectively, of variable *A* for *m* samples. Because RReliefF considers collinearity among the predictor variables, it has an advantage over other feature selection methods that are solely based on statistical measures (e.g. correlation coefficient and signal to noise ratio; Son et al., 2015).

2.5.2. RF regression

RF regression was implemented to build prediction models for AGB and yield using VIs and crop height. RF regression is a supervised machine learning algorithm that combines a large number of regression trees (*ntree*), each consisting of a random subset of one third of the predictor variables (Wang et al., 2016). The *ntree* value was selected by optimising the root mean square error of calibration. RF regression was performed as follows:

(a) Bootstrapped samples were randomly selected from the original calibration dataset containing approximately two thirds of the randomly selected input variables. The remainder of the samples were referred to as out-of-bag (OOB) samples.

(b) Following modifications on each node, each regression tree was independently trained on a bootstrapped subset iteratively with one third of the variables randomly selected until the forest is grown to *ntree*.

(c) For each bootstrapped iteration, the OOB data can be predicted by fitting the variable vector to the trees. The predictions from each tree in the forest were then aggregated by taking the mean of all trees. The OOB error was calculated following comparisons with ground truth data.

RF regression is not sensitive to collinearity among variables, ensuring prediction accuracy and reducing overfitting (Moisen, 2008). To optimise model calibration, the number of trees is determined when there is no noticeable improvement in prediction accuracy with increased trees. An independent dataset was then used to validate the accuracy and robustness of the RF model. Root mean squares errors for prediction (RMSEP) and residual prediction deviation (RPD; Valente et al., 2013), defined as the ratio of the standard deviation of the reference values in the training dataset to RMSEP, were used to assess model accuracy and robustness. RPD values were classified based on the published criteria (Yang, 2011): (1) the model is not applicable if the RPD is < 1.5; (2) the model can only discriminate between low and high value groups if the RPD is 1.5–2; (3) the model can perform coarse quantitative prediction if the RPD is 2–2.5; and (4) the model can perform prediction accurately if RPD is > 2.5.

2.5.3. Partial least squares regression

By splitting the spectral data into calibration and test datasets, PLS regression analysis was used to developed multiple prediction models to estimate the mathematical relationship between a set of independent (*X* matrix; N_{sample_num} x $K_{variable\_num}$) and dependent variables (*Y* matrix; N_{sample_num} x 1) including crop height, AGB, and yield. PLS regression decomposes both the dependent (*Y*) and independent (*X*) variables into a number of principal components, and can accommodate highly correlated variables and over-fitting. The PLS regression model applies the component projection to find the latent structure of a dataset. By selecting the optimal number of latent variables (LVs), the regression variables can be reduced from all wavelengths with heavy collinearity to a few independent principal components and transformed into scores. The prediction model can be described using Eq. 4, and the regression coefficients *B* can be calculated by regressing *Y* onto the wavelength scores T_{LVs} as shown in Eq. 5:

$$\bar{Y} = X * B + E = X * W_{LVs}^{*} * C + E = T_{LVs} * C + E \tag{4}$$

$$W_{LVs}^{*} = W_{LVs} * \left(P^{'} * W_{LVs}\right)^{-1} \tag{5}$$

where $\bar{Y}$ represents the estimated dependent variables, *X* represents the predictor variables, *B* represents the regression coefficients, E is the residual error matrix, W_{LVs} represents a set of orthogonal projection axes called PLS weights, T_{LVs} is the score matrix determined using the PLS algorithm, and *P* and *C* are the loadings of *X* and *Y*, respectively.

Leave-one-out cross validation (LOOCV) was used to determine the optimal number of LVs with the optimal coefficient of determination for cross validation (r_v^2) and minimum root mean squares errors for cross validation (RMSECV). A test dataset was used to validate the accuracy and robustness of the derived PLS model using the coefficient of determination for prediction (r_p^2), RMSEP, and RPD as the criteria for assessing model performance (Li et al., 2018).

2.5.4. Crop height estimation

Crop height can be estimated using either a DSM generated from the 3D model of the UAV imaging (Bendig et al., 2014) or by modelling the spectra extracted from UAV hyperspectral imaging (Capolupo et al., 2015). The DSM model generated from the 3D reconstruction of UAV-based RGB imagery is in a TIF image format, and the 16 bit float intensity value of each pixel represents the absolute height of the object in the pixel. The digital elevation model (DEM) that represents the absolute elevation of the bare ground under the canopy was estimated by interpolating values extracted from the neighbouring bare soil buffer zones between plots, performed using ESRI ArcGIS 10.2.2 and the ordinary Kriging method (Geipel et al., 2014; Mathews and Jensen, 2013). Crop height was then estimated as the difference between the DSM and DEM as follows:

$$nDSM = DSM - DEM \tag{6}$$

where nDSM is the estimated absolute plant height. Because crop height was measured between the ground and the top of the canopy, local maxima (high intensity pixels surrounded by lower intensity pixels), were applied to identify the top of the canopy in the nDSM (Garrido et al., 1998). Convolution with a sliding window was applied to the entire nDSM image so that the maxima could be identified for each window, and the average value of local maxima was used to indicate the crop height in each sampling plot. The DSM and DEM models were applied to the 60 plots of Experiment 1 (N fertiliser input) due to the large buffer zone at this site, and ground-truth data measured at 90 DAP were used to validate model performance. PLS regression of crop height with the full spectra extracted from UAV-based hyperspectral imagery permits the direct estimation of crop height without a DEM (Capolupo et al., 2015). The average spectra from Experiments 2 and 3 at both 60 and 90 DAP were used for model calibration (n = 168). As with the nDSM-based method, the 60 spectra extracted from Experiment 1 at 90 DAP were used as a test dataset.

2.5.5. Biomass estimation and yield prediction models

Both PLS and RF regression models were constructed to estimate AGB and predict yield. The PLS regression model was developed with predictor variables based on the full wavelength range. Both narrow-band VIs and estimated crop height data were used to develop the RF regression model while only VIs were used as predictor variables for yield prediction. The average spectra extracted from Experiment 1 at both 60 and 90 DAP were used for the development of the biomass estimation models (n = 120), and the total spectra were split into training and test datasets with a split ratio of 75:25. Separate yield prediction models were developed for the two flight surveys. Because ground-based yield data for five plots were not recorded, the remaining 139 average spectra values were divided into training and test datasets with a split ratio of 75:25. The training spectra for both AGB and yield predictions were randomly selected to maximise the data range of the training dataset.

3. Results

3.1. Ground truth data

The minimum, maximum, mean, and standard deviation of dry and fresh AGB and yield data are shown in Table 2. The large range of data ensures the robustness of the models derived from the data. Dry and fresh AGB were highly correlated with each other (r^2 = 0.94). The correlation of dry/fresh AGB with yield was not calculated because the spectra were taken from different plots.

Table 2. Statistics of ground-truth data for dry and fresh potato above ground biomass (AGB) and yield for model calibration and test datasets.

Parameters	Calibration				Prediction			
	Min	Max	Mean	SD	Min	Max	Mean	SD
Dry AGB (ton ha^{-1})	0.74	9.04	2.85	2.08	0.55	7.04	3.28	2.04
Fresh AGB (ton ha^{-1})	2.38	58.36	14.99	13.13	4.02	43.05	17.97	11.93
Yield (ton ha^{-1})	1.14	5.84	3.01	0.97	1.22	4.85	2.89	0.97

Table 3. Summary of the prediction models used in the study

Parameter		Models	Variables	Model index
Crop height		Linear regression	nDSM	CH1
		PLSR	Full wavelength	CH2
Fresh AGB		RF	CI1, Crop height and MSR	FA1
		PLSR	Full wavelength	FA2
Dry AGB		RF	CI1, Crop height and MSR	DA1
		PLSR	Full wavelength	DA2
Yield	60 DAP	RF	CI1, MCARI, height	Y160
		PLSR	Full wavelength	Y260
	90 DAP	RF	CI1, MCARI, height, Ratio2, CI2	Y190
		PLSR	Full wavelength	Y290

3.2. Crop height estimation

The DEM (Fig. 3b) was used to estimate the elevation of bare soil by interpolating the elevation values from neighbouring buffer zones in the DSM (Fig. 3a), and the resulting nDSM image (Fig. 3c), representing crop heights, is shown in Fig. 3c. Crop height estimated using the nDSM with local maxima (Fig. 4d) and the PLS regression model (Fig. 4d) are compared with ground truth data in Experiment 1 at 90 DAP. The nDSM-derived crop heights show a high correlation with the ground truth data (CH1, Table 3, r_p^2 = 0.93, RMSEP = 6.39 cm) and the RPD value (2.89) indicates robust model prediction. The PLS regression model with full wavelength variables also performed reasonably well (CH2, Table 3, r_p^2 = 0.85, RMSEP = 7.24 cm, RPD = 2.55), although worse than nDSM method (Fig. 4e). The PLS regression model statistics are shown in Table 4. The nDSM model performed better than the PLS regression method, and because the impact of cultivar, illumination and canopy density on the PLS crop height model was not adequately investigated in the preliminary study. Hence, the crop height estimated using the nDSM model was applied for AGB estimation.

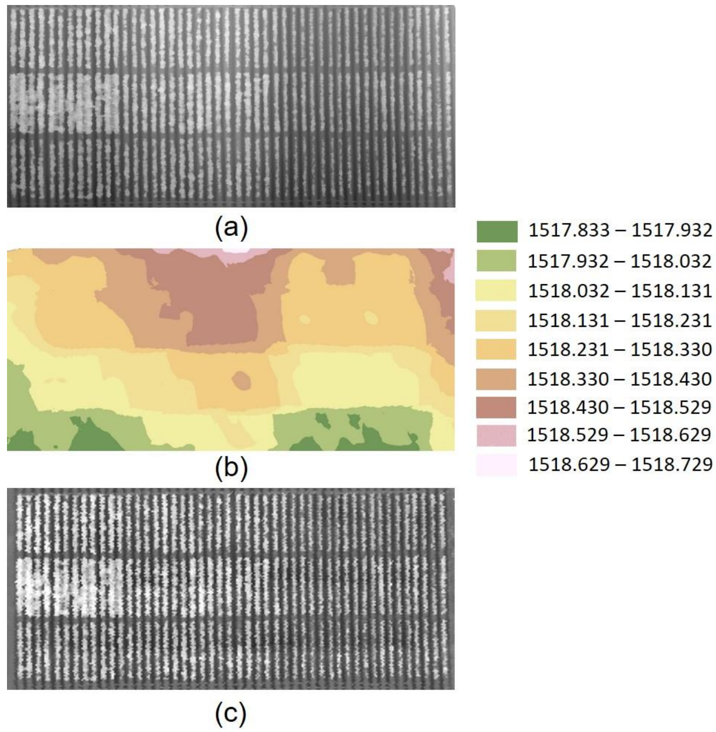


Figure 3. Sample images of the original digital surface model (DSM) (a), estimated digital elevation model (DEM) with hot map and elevation scale (b), and the resulting nDSM, representing crop height (c).

Table 4. Calibration, leave-one-out cross validation (LOOCV), and independent prediction statistics of the partial least squares (PLS) regression model for crop height estimation.

Parameter	LVs	r_c^2	RMSEC (cm)	r_v^2	RMSECV (cm)	r_p^2	RMSEP (cm)	RPD
Crop height	9	0.90	4.89	0.87	5.71	0.85	7.24	2.55

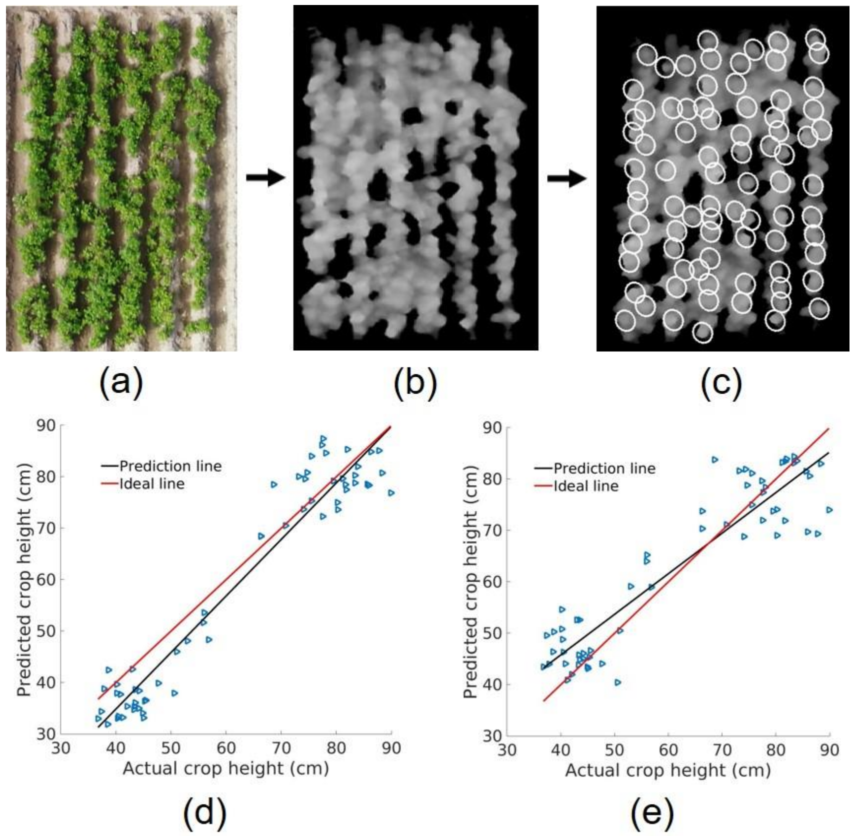


Figure 4. Original RGB image for a single example plot (a), the resulting nDSM image (b) and the nDSM image with local maxima labelled (c). Comparison of crop heights estimated using nDSM (d) and PLS regression (e) with ground-based manual assessment.

3.3. Estimation of AGB using the RF and PLS regression models

3.3.1. RF regression model

The estimated crop heights using the nDSM method in Experiment 1 were used as predictors in the RF regression models for AGB estimation from both flight surveys. The importance of all predictors (VIs and crop heights) was evaluated using RReliefF (Figs. 5b and d). The best prediction accuracy for both dry and fresh AGB was achieved using only three predictors; CI1 (Table 1), crop height, and MSR (Table 1). No apparent change in the OOB error was observed when the number of trees reached approximately 300; hence, this value was used as *ntree* in the RF model. The prediction results for the test dataset showed that the RF models can estimate both fresh (FA1, Table 3, r_p^2 = 0.90, RMSEP = 3.71 ton ha^{-1}, RPD = 3.22) and dry (DA1, Table 3, r_p^2 = 0.92, RMSEP = 0.57 ton ha^{-1}, RPD = 3.55) AGB accurately (Figs. 5a and c). Both models showed decreased prediction accuracy when the AGB was high, probably due to saturation of the spectral indices at high vegetation densities (Maimaitijiang et al., 2019).

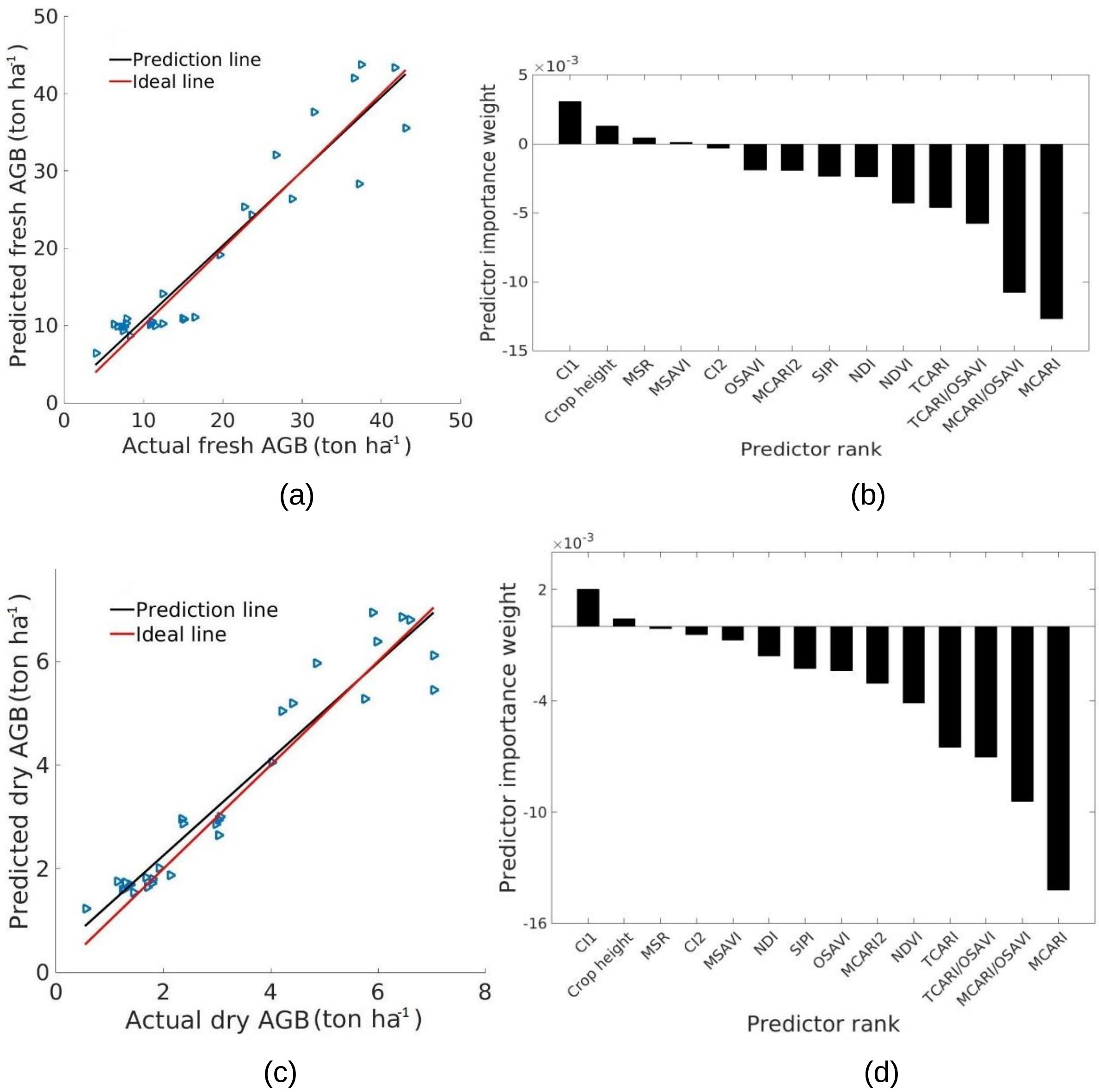
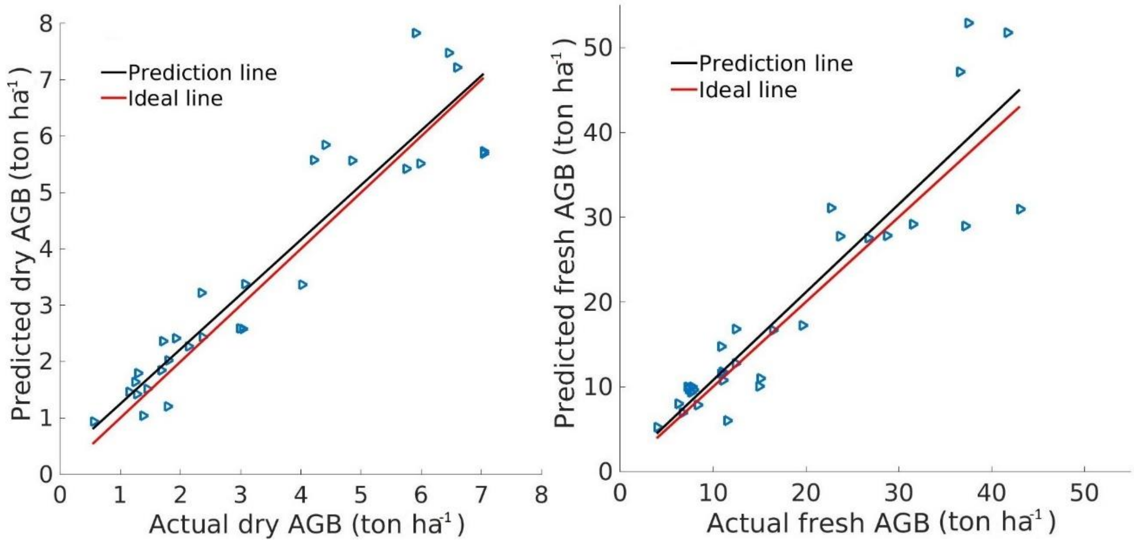


Figure 5. Prediction of fresh AGB using the random forest (RF) regression model (a) and the importance of all predictor variables (VIs and crop height) (b). Prediction of dry AGB using the RF regression model (c) and the importance of all predictor variables (d).

3.3.2. PLS regression model

PLS regression models were developed with full wavelength variables to estimate fresh (FA2, Table 3) and dry (DA2, Table 3) AGB. Results show that the prediction accuracy is higher for dry AGB compared to fresh AGB (Fig. 6 and Table 5) with an RPD > 2.5. The overall performance of the PLS regression models was slightly worse compared to the RF regression models. The deviation between actual and predicted values is larger than for the RF regression models, indicating that plant height is significant for AGB estimation, especially for high canopy densities.



(a) (b)

Figure 6. Dry (a) and fresh (b) AGB prediction using the partial least squares (PLS) regression model.

Table 5. Calibration, LOOCV, and independent prediction statistics using the PLS regression model for AGB estimation.

Parameter	LVs	r_c^2	RMSEC (ton ha^{-1})	r_v^2	RMSECV (ton ha^{-1})	r_p^2	RMSEP (ton ha^{-1})	RPD
Fresh AGB	10	0.85	4.87	0.78	5.99	0.83	5.47	2.18
Dry AGB	8	0.88	0.72	0.82	0.88	0.88	0.88	2.68

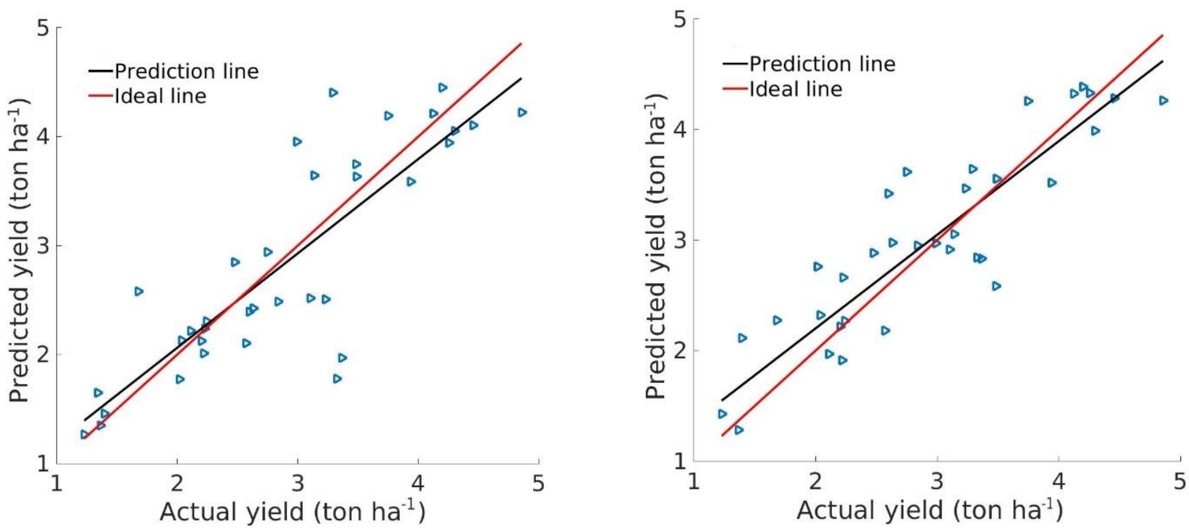
3.4. Yield prediction

3.4.1. RF regression model

Separate RF regression models were constructed based on the VIs and crop height values derived from the two flight surveys. Both models showed the best prediction using the optimal predictors selected by the RReliefF algorithm. Due to insufficient unplanted buffer zones in Experiments 2 and 3, the nDSM method could not be used for crop height estimation. Therefore, manually measured crop height values were used to validate the impact of crop height on yield prediction for these experiments. The 60 DAP model showed insufficient yield prediction accuracy (Y160, Table 3, r_p^2 = 0.44, RMSEP = 0.73 ton ha^{-1}, RPD = 1.34) with the predictive variables CI1, MCARI, and crop height demonstrating the best performance. The 90 DAP model performed better (Y190, Table 3, r_p^2 = 0.63, RMSEP = 0.63 ton ha^{-1}, RPD = 1.55) using crop height and four VIs (CI1, MCARI, MCARI/OSAVI, and CI2) as predictors (Fig. 5). However, the RPD indicates that the model can only discriminate between low and high yield values rather than providing accurate yield prediction.

3.4.2. PLS regression model

The full spectra 90 DAP PLS regression model (Y290, Table 3) showed markedly improved predictive skill compared to the 60 DAP model (Y260, Table 3; Fig. 7). The r_p^2 and RPD values (Table 6) indicate the feasibility of using the full spectra PLS regression model for coarse yield prediction.



(a) (b)

Figure 7. Yield prediction using the PLS regression model based on the spectra taken 60 (a) and 90 (b) days after planting (DAP)

Table 6. Calibration, LOOCV, and independent prediction statistics of the PLS regression model for yield prediction.

Date	LVs	r_c^2	RMSEC (ton ha^{-1})	r_v^2	RMSECV (ton ha^{-1})	r_p^2	RMSEP (ton ha^{-1})	RPD
60 DAP	6	0.80	0.42	0.77	0.46	0.69	0.56	1.75
90 DAP	11	0.80	0.43	0.66	0.57	0.81	0.42	2.29

4. Discussion

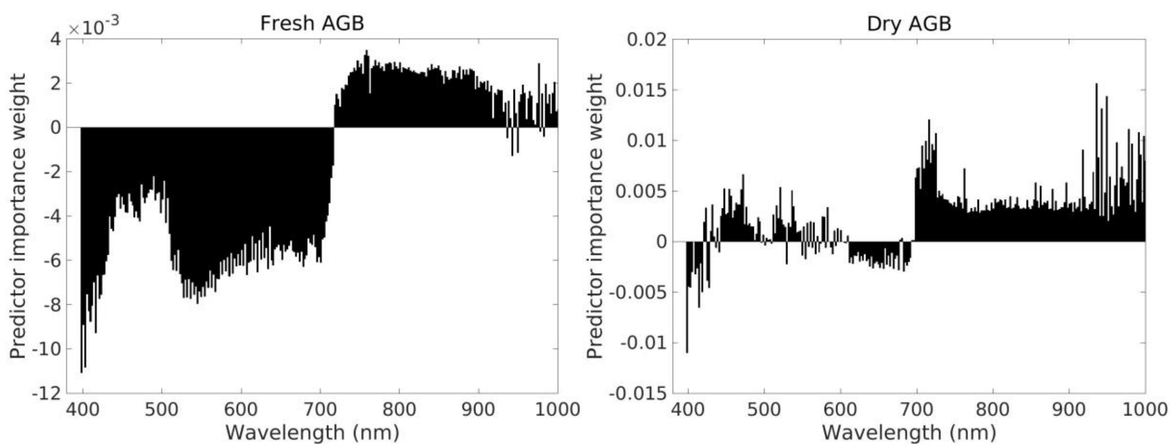
Limited research is available regarding the prediction of AGB and yield for potato crops using remote sensing techniques. Previous potato crop studies used either a single (Millard et al., 1990) or unnamed cultivar (Al-Gaadi et al., 2016; Morier et al., 2015). They found that model performance based on a single broad-band VI with potato AGB varied with N fertiliser treatment (Millard et al., 1990), and was insufficient for yield prediction (Al-Gaadi et al., 2016). Low-altitude UAVs with a hyperspectral imaging sensor, as used in the present study, provide a high spatial and spectral resolution. Six potato cultivars were planted under different treatments of N, K, and mixed organic-inorganic compound fertilisers, providing varied AGB and yield data and ensuring the robustness of the derived models. Reflectance spectra can be impacted by illuminations; however, using multiple VIs can reduce this effect by calculating the relative difference or ration among wavelengths. Furthermore, flight surveys were carried out on two occasions under different light conditions, further improving the robustness of the crop height and AGB estimations.

Manual assessment of crop heights is time consuming; thus, only a small portion of crops can be measured, leading to inaccuracies. As a high-throughput phenotyping method, UAV-based imaging was introduced to estimate potato canopy height. The nDSM-based method provides a low-cost solution compared to hyperspectral imaging techniques; however, interpolation for DEM estimation requires a large unplanted buffer zone within the target field, which is not always practical in commercial farming. In this study, the nDSM method could not be used to estimate crop heights in Experiments 2 and 3 because of the lack of a buffer zone. Alternatively, the DEM may be obtained by imaging the bare ground with a UAV before crop emergence, which should be applied in further study. Although this requires a sufficient number of ground control points and measurements using Real Time Kinetic Global Navigation Satellite System equipment (Geipel et al., 2014), it is more practical for commercial farms and flexible for different experimental designs. In previous studies, only the mean, standard deviation, and maximum and minimum crop height could be measured automatically using the nDSM method, while individual crop heights still required manual extraction (Han et al., 2019; Holman et al., 2016; Tilly et al., 2015). The local maxima algorithm enables the automated identification of the maximum height of individual plants in the nDSM image and is more accurate than averaging the nDSM image, which invariably leads to underestimation (Aasen et al., 2015; Han et al., 2019).

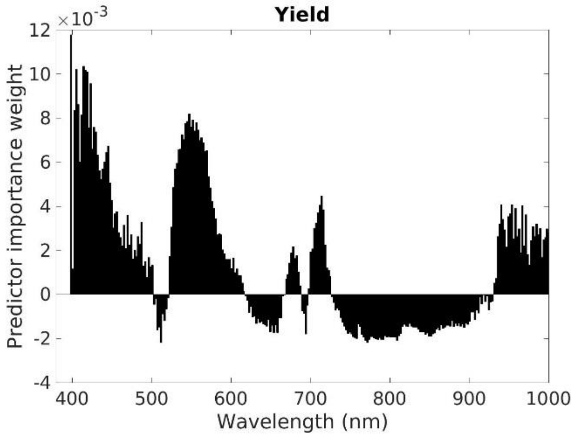
RF regression was successfully applied to AGB estimation using VIs as predictor variables and performed better than MLR, SVM, and ANN (Han et al., 2019; Wang et al., 2016). Our results show that both RF and PLS regression models demonstrate satisfactory prediction accuracy for AGB. A combination of two VIs (CI1 and MSR) and crop height were identified as the key predictors by RReliefF. Consistent with previous studies (Freeman et al., 2007; Tilly et al., 2015), crop height was highly correlated with AGB and the inclusion of crop height with VIs improved the accuracy of the AGB prediction. However, importance of crop height as a predictor was lower than CI1, which can most likely be attributed to the multiple varieties of potato used in this experiment. The canopy morphology of different potato varieties differ. For example, *Favorita* has a lower, more widespread canopy compared with Z18. Similar conclusion can be found in the study of Bendig et al. (2014) that cultivar difference such as lodging and non-lodging is one constraint for biomass prediction of barley by crop height. Furthermore, both early maturing (*Favorita*, Z5 and Z10) and late maturing varieties (Z18, Z19 and *Shepody*) were used in this study. When the potato plant grows to a certain height, the canopy will continue to grow and spread. For late maturing cultivars, it is possible that the canopy was not yet well developed, despite reaching its maximum height. This conclusion is also consistent with the study of Bendig et al. (2014), which showed the cultivar difference is one constraint for biomass prediction by crop height. Predictive skill was lower for fresh as compared to dry AGB. This was probably due to the varied weather conditions on flight survey days resulting in different water contents in the fresh AGB. However, this would not impact the estimation of plant height (Tilly et al., 2015). CI1 was the most important VI for estimating AGB in this study. This index is not directly related to AGB (Tilly et al., 2015); however, it shows good correlation with chlorophyll, N content, and leaf area index (Clevers et al., 2012; Gitelson et al., 2003), which are all related to AGB (Babar et al., 2006; Holben et al., 1980). The PLS regression models based on full wavelengths performed worse than the RF regression models. We attribute this to the lack of crop height information and redundancy in some wavelengths. The application of VIs with selected wavelengths rather than a full spectra for AGB estimation can also facilitate the conversion from hyperspectral to multispectral cameras using selected bands, leading to a potential reduction in camera cost.

Yield prediction models using VIs as predictors showed insufficient accuracy for both flight surveys, although the accuracy was still greater than those obtained in previous studies using a single VI (Al-Gaadi et al., 2016; Morier et al., 2015). PLS regression models based on the full wavelength spectra make full use of the rich spectral information from hyperspectral imaging data, overcoming the limitations of using a few selected spectra. Similar conclusion was also found in the study of Montesinos-López et al. (2017), which showed using statistical models with all bands simultaneously increased the prediction accuracy more than using VIs along. When RReliefF analysis was applied to assess the importance of each individual wavelength (Mahlein et al., 2013), most of the key wavelengths for both fresh and dry AGB estimation were within the near infrared region (Figs. 8a and b), explaining why near infrared VIs could predict biomass with good accuracy. Yield is affected by many factors and its prediction can be more complicated as compared to AGB. Figure 8c shows that the key wavelengths for yield prediction are located across the visible and near infrared range, except for the red-edge region, illustrating why the VIs selected in this study were not adequate for yield prediction.

Further study is required to include more VIs within the visible region to improve current RF regression models. The inclusion of crop height resulted in improved model accuracy for yield prediction. However, it should be noted that the lack of unplanted buffer zones in the K and mixed compound fertiliser experiments meant that only manually observed data were evaluated. Crop heights derived from nDSM should be incorporated into prediction models because they are likely to be more accurate than manually estimated crop heights from limited sampling. Further studies will also investigate the significance of the volume metric derived from the multiplication of the plant height and the area covered by the plant for both AGB and yield prediction, which were successfully used for estimating the AGB of soybean and maize (Han et al., 2019; Maimaitijiang et al., 2019). Prediction accuracy of the PLS regression model at 90 DAP is satisfactory in this study; however, further research is needed to understand the relationship between prediction accuracy and survey timing relative to crop development. Both AGB and yield estimation models were investigated and developed under similar sowing density across all plots. Future study is essential to design the experiments and introduce sowing density as variable to understand its impact.



(a) (b)



(c)

Figure 8. Predictor importance of each individual wavelength for the prediction of fresh (a) and dry (b) AGB and yield (c).

5. Conclusion

This study used a low altitude UAV equipped with RGB and hyperspectral imaging sensors to predict potato biomass and yield. Multiple VIs derived from hyperspectral imaging data and plant heights measured using an nDSM-based method were used as predictor variables in RF and PLS modelling. CI1, crop height, and MSR were selected as the most important predictors. In terms of AGB, the RF regression model had better prediction accuracy compared to the PLS regression model based on the full spectra. Conversely, the PLS regression model performed better than the RF regression model in predicting potato yield. Yield prediction using survey data one month prior to harvesting was satisfactory.

We conclude that UAV-based hyperspectral imaging is a promising remote sensing technique for predicting potato AGB and yield, and can be adopted for site-specific crop management.

Acknowledgements

This article was supported by grants from The National Key Research and Development Program of China (2018YFD0200803), China Agriculture Research System (CARS-09-P12) awarded to Chunsong Bian, Breeding new varieties for advantageous agricultural industries in Ningxia-Digital breeding system for potato (2019NYYZ01-4) awarded to Dr. Jiangang Liu and Agri-Tech in China: Newton Network + (LG005) awarded to Dr. Bo Li.

Appendix A. Supplementary data

Treatment*	F1	F2	F3	F4
	Compound fertilizer (CF, kg ha^{-1}) ($N:P_2O_5:K_2O$ =15:15:15)	F1+Soil Conservation fertilizer (SCF, kg ha^{-1})	F1+Soil Conservation fertilizer (SCF, kg ha^{-1})	F1+Organic -inorganic fertilizer (OIF, kg ha^{-1})
Base Fertilizer	CF300	CF300+SCF300	CF300+SCF150	CF300+OIF300
Treatment*	F5	F6	F7	F8
	F1+Organic-inorganic fertilizer (OIF, kg ha^{-1})	F1+Compound microorganism (CM, kg ha^{-1})	F1+Compound microorganism (CM, kg ha^{-1})	F1+25%F1
Base Fertilizer	CF300+OIF150	CF300+CM80	CF300+CM160	CF600

*CF: Sino-Arab Chemical Fertilizers Co. Ltd. (SACF), $N:P_2O_5:K_2O$ = 15:15:15; SCF: Guizhou Bao Tu Ecological Recycling Agriculture Technology Co. Ltd., N:P:K = 6:4:10; OIF: Yunnan Tumama Fertilizers Co.,Ltd, $N:P_2O_5:K_2O$ = 8:8:14, Organic matter ≥ 12%; CM: *Bacillus subtilis* / *Bacillus licheniformis*, complex fermentation, microbial propagules ≥ 0.2 billion per gram.

Conflicts of interest

The authors declare no conflict of interest

References

Aasen, H., Burkart, A., Bolten, A., Bareth, G., 2015. Generating 3D hyperspectral information with lightweight UAV snapshot cameras for vegetation monitoring: From camera calibration to quality assurance. ISPRS J. Photogramm. Remote Sens. 108, 245-259. https://doi.org/10.1016/j.isprsjprs.2015.08.002

Al-Gaadi, K.A., Hassaballa, A.A., Tola, E., Kayad, A.G., Madugundu, R., Alblewi, B., Assiri, F., 2016. Prediction of Potato Crop Yield Using Precision Agriculture Techniques. PLoS One 11, e0162219.

Aparicio, N., Villegas, D., Casadesus, J., Araus, J.L., Royo, C., 2000. Spectral Vegetation Indices as Nondestructive Tools for Determining Durum Wheat Yield. Agron J. 91, 83. https://doi.org/10.2134/agronj2000.92183x

Babar, M.A., Reynolds, M.P., van Ginkel, M., Klatt, A.R., Raun, W.R., Stone, M.L., 2006. Spectral Reflectance to Estimate Genetic Variation for In-Season Biomass, Leaf Chlorophyll, and Canopy Temperature in Wheat. Crop Sci. 46, 1046. https://doi.org/10.2135/cropsci2005.0211

Bai, J.-H., Jun-hua, B.A.I., Shao-kun, L.I., Wang, K.-R., Xue-yan, S.U.I., Chen, B., Wang, F.-Y., 2007. Estimating Aboveground Fresh Biomass of Different Cotton Canopy Types with Homogeneity Models Based on Hyper Spectrum Parameters. Agricultural Sciences in China. 6, 437-445. https://doi.org/10.1016/s1671-2927(07)60067-4

Ballesteros, R., Ortega, J.F., Hernandez, D., Moreno, M.A., 2018. Onion biomass monitoring using UAV-based RGB imaging. Precis Agric. 840-857. https://doi.org/10.1007/s11119-018-9560-y

Bendig, J., Bolten, A., Bennertz, S., Broscheit, J., Eichfuss, S., Bareth, G., 2014. Estimating Biomass of Barley Using Crop Surface Models (CSMs) Derived from UAV-Based RGB Imaging. Remote Sens. 6, 10395-10412. https://doi.org/10.3390/rs61110395

Cao, X., Luo, Y., Zhou, Y., Fan, J., Xu, X., West, J.S., Duan, X., Cheng, D., 2015. Detection of powdery mildew in two winter wheat plant densities and prediction of grain yield using canopy hyperspectral reflectance. PLoS One 10, e0121462.

Capolupo, A., Kooistra, L., Berendonk, C., Boccia, L., Suomalainen, J., 2015. Estimating Plant Traits of Grasslands from UAV-Acquired Hyperspectral Images: A Comparison of Statistical Approaches. ISPRS J. Photogramm. Remote Sens. 4, 2792-2820.

https://doi.org/10.3390/ijgi4042792

Chen, J.M., 1996. Evaluation of vegetation indices and a Modified Simple Ratio for boreal applications. https://doi.org/10.4095/218303

Clark, M.L., Roberts, D.A., Ewel, J.J., Clark, D.B., 2011. Estimation of tropical rain forest aboveground biomass with small-footprint lidar and hyperspectral sensors. Remote Sens. Environ. 115, 2931-2942. https://doi.org/10.1016/j.rse.2010.08.029

Clevers, J.G.P.W., Jan G P, Kooistra, L., 2012. Using Hyperspectral Remote Sensing Data for Retrieving Canopy Chlorophyll and Nitrogen Content. IEEE J Sel Top Appl Earth Obs Remote Sens. 5, 574-583. https://doi.org/10.1109/jstars.2011.2176468

Clevers, J.G.P.W., J G P, van der Heijden, G.W.A.M., Verzakov, S., Schaepman, M.E., 2007. Estimating Grassland Biomass Using SVM Band Shaving of Hyperspectral Data. Photogramm Eng Remote Sensing. 73, 1141-1148. https://doi.org/10.14358/pers.73.10.1141Colomina, I., Molina, P., 2014.

Colomina, I., Molina, P., 2014. Unmanned aerial systems for photogrammetry and remote sensing: A review. ISPRS J. Photogramm. Remote Sens. 92, 79–97. https://doi.org/10.1016/j.isprsjprs.2014.02.013

Datt, B., 1999. A New Reflectance Index for Remote Sensing of Chlorophyll Content in Higher Plants: Tests using Eucalyptus Leaves. J Plant Physiol. 154, 30-36. https://doi.org/10.1016/s0176-1617(99)80314-9

Daughtry, C., 2000. Estimating Corn Leaf Chlorophyll Concentration from Leaf and Canopy Reflectance. Remote Sens. Environ. 74, 229-239. https://doi.org/10.1016/s0034-4257(00)00113-9

Domingues Franceschini, M.H., Bartholomeus, H., van Apeldoorn, D., Suomalainen, J., Kooistra, L., 2017. Correction: Domingues Franceschini, M.H.; et al. Intercomparison of Unmanned Aerial Vehicle and Ground-Based Narrow Band Spectrometers Applied to Crop Trait Monitoring in Organic Potato Production. Sensors. 17, 1428. https://doi.org/10.3390/s17102265

Freeman, K.W., Girma, K., Arnall, D.B., Mullen, R.W., Martin, K.L., Teal, R.K., Raun, W.R., 2007. By-Plant Prediction of Corn Forage Biomass and Nitrogen Uptake at Various Growth Stages Using Remote Sensing and Plant Height. Agron J. 99, 530. https://doi.org/10.2134/agronj2006.0135

Garrido, L., Salembier, P., Garcia, D., 1998. Extensive operators in partition lattices for image sequence analysis. Signal Processing. 66, 157-180. https://doi.org/10.1016/s0165-1684(98)00004-8

Geipel, J., Link, J., Claupein, W., 2014. Combined Spectral and Spatial Modeling of Corn Yield Based on Aerial Images and Crop Surface Models Acquired with an Unmanned Aircraft System. Remote Sens. 6, 10335-10355. https://doi.org/10.3390/rs61110335

Gitelson, A.A., Merzlyak, M.N., 1996. Signature Analysis of Leaf Reflectance Spectra: Algorithm Development for Remote Sensing of Chlorophyll. J Plant Physiol. 148, 494-500. https://doi.org/10.1016/s0176-1617(96)80284-7

Gitelson, A.A., Viña, A., Arkebauer, T.J., Rundquist, D.C., Keydan, G., Leavitt, B., 2003. Remote estimation of leaf area index and green leaf biomass in maize canopies. Geophys Res Lett. 30. https://doi.org/10.1029/2002gl016450

Greaves, H.E., Vierling, L.A., Eitel, J.U.H., Boelman, N.T., Magney, T.S., Prager, C.M., Griffin, K.L., 2015. Estimating aboveground biomass and leaf area of low-stature Arctic shrubs with terrestrial LiDAR. Remote Sens. Environ. 164, 26–35. https://doi.org/10.1016/j.rse.2015.02.023

Haboudane, D., 2004. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of Precis Agric. Remote Sens. Environ. 90, 337-352. https://doi.org/10.1016/j.rse.2003.12.013

Haboudane, D., Miller, J.R., Tremblay, N., Zarco-Tejada, P.J., Dextraze, L., 2002. Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. Remote Sens. Environ. 81, 2-3. https://doi.org/10.1016/s0034-4257(02)00018-4

Han, L., Yang, G., Dai, H., Xu, B., Yang, H., Feng, H., Li, Z., Yang, X., 2019. Modeling maize above-ground biomass based on machine learning approaches using UAV remote-sensing data. Plant Methods 15, 10.

Hansen, P.M., Jørgensen, J.R., Thomsen, A., 2002. Predicting grain yield and protein content in winter wheat and spring barley using repeated canopy reflectance measurements and partial least squares regression. J Agric Sci. 139. https://doi.org/10.1017/s0021859602002320

Hansen, P.M., Schjoerring, J.K., 2003. Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. Remote Sens. Environ. 86, 842-553. https://doi.org/10.1016/s0034-4257(03)00131-7

Herrmann, I., Karnieli, A., Bonfil, D.J., Cohen, Y., Alchanatis, V., 2010. SWIR-based spectral indices for assessing nitrogen content in potato fields. Int. J. Remote Sens. 31, 5127-5143. https://doi.org/10.1080/01431160903283892

Holben, Brent N., Compton J. Tucker, and Cheng-Jeng Fan, 1980. spectral assessment of soybean leaf area and leaf biomass. Photogramm. Eng. Remote Sens. 46, 651–656.

Holman, F., Riche, A., Michalski, A., Castle, M., Wooster, M., Hawkesford, M., 2016. High Throughput Field Phenotyping of Wheat Plant Height and Growth Rate in Field Plot Trials Using UAV Based Remote Sensing. Remote Sens. 8, 1031. https://doi.org/10.3390/rs8121031

Huang, J., Wang, X., Li, X., Tian, H., Pan, Z., 2013. Remotely Sensed Rice Yield Prediction Using Multi-Temporal NDVI Data Derived from NOAA's-AVHRR. PLoS ONE. https://doi.org/10.1371/journal.pone.0070816

Jain, N., Ray, S.S., Singh, J.P., Panigrahy, S., 2007. Use of hyperspectral data to assess the effects of different nitrogen applications on a potato crop. Precis. Agric. 8, 225-239. https://doi.org/10.1007/s11119-007-9042-0

Kira, K., Rendell, L.A., 1992. A Practical Approach to Feature Selection. Machine Learning Proceedings 1992. 249-256. https://doi.org/10.1016/b978-1-55860-247-2.50037-1

Li, B., Cobo-Medina, M., Lecourt, J., Harrison, N., Harrison, R.J., Cross, J.V., 2018. Application of hyperspectral imaging for nondestructive measurement of plum quality attributes. Postharvest Biol. Technol. 141, 8-15. https://doi.org/10.1016/j.postharvbio.2018.03.008

Li, B., Xu, X., Han, J., Zhang, L., Bian, C., Jin, L., Liu, J., 2019. The estimation of crop emergence in potatoes by UAV RGB imagery. Plant Methods. 15, 15.

Li, F., Mistele, B., Hu, Y., Yue, X., Yue, S., Miao, Y., Chen, X., Cui, Z., Meng, Q., Schmidhalter, U., 2012. Remotely estimating aerial N status of phenologically differing winter wheat cultivars grown in contrasting climatic and geographic zones in China and Germany. Field Crops Res. 138, 21-32. https://doi.org/10.1016/j.fcr.2012.09.002

Mahlein, A.-K., -K. Mahlein, A., Rumpf, T., Welke, P., -W. Dehne, H., Plümer, L., Steiner, U., -C. Oerke, E., 2013. Development of spectral indices for detecting and identifying plant diseases. Remote Sens. Environ, 128, 21-30. https://doi.org/10.1016/j.rse.2012.09.019

Maimaitijiang, M., Sagan, V., Sidike, P., Maimaitiyiming, M., Hartling, S., Peterson, K.T., Maw, M.J.W., Shakoor, N., Mockler, T., Fritschi, F.B., 2019. Vegetation Index Weighted Canopy Volume Model (CVM VI) for soybean biomass estimation from Unmanned Aerial System-based RGB imagery. ISPRS J. Photogramm. Remote Sens. 151, 27–41. https://doi.org/10.1016/j.isprsjprs.2019.03.003

Mathews, A.J., Jensen, J.L.R., 2013. Visualizing and quantifying vineyard canopy LAI using an unmanned aerial vehicle (UAV) collected high density structure from motion point cloud. Remote Sens. 5, 2164–2183. https://doi.org/10.3390/rs5052164

Millard, P., Wright, G.G., Adams, M.J., Birnie, R.V., Whitworth, P., 1990. Estimation of light interception and biomass of the potato (Solanum tuberosum L.) from reflection in the red and near-infrared spectral bands. Agric For Meteorol. 53, 19-31. https://doi.org/10.1016/0168-1923(90)90121-l

Moisen, G.G., 2008. Classification and Regression Trees. Encyclopedia of Ecology. 582-588. https://doi.org/10.1016/b978-008045405-4.00149-x

Montesinos-López, O.A., Montesinos-López, A., Crossa, J., de Los Campos, G., Alvarado, G., Suchismita, M., Rutkoski, J., González-Pérez, L., Burgueño, J., 2017. Predicting grain yield using canopy hyperspectral reflectance in wheat breeding data. Plant Methods 13, 4.

Morier, T., Cambouris, A.N., Chokmani, K., 2015. In-Season Nitrogen Status Assessment and Yield Estimation Using Hyperspectral Vegetation Indices in a Potato Crop. 107, 1295. Agron J. https://doi.org/10.2134/agronj14.0402

Mutanga, O., Skidmore, A.K., 2004. Narrow band vegetation indices overcome the saturation problem in biomass estimation. Int. J. Remote Sens. 25, 3999-4014. https://doi.org/10.1080/01431160310001654923

Ngo-Ye, T.L., Sinha, A.P., 2012. Analyzing Online Review Helpfulness Using a Regressional ReliefF-Enhanced Text Mining Method. ACM Transactions on Management Information Systems. 3, 1-20. https://doi.org/10.1145/2229156.2229158

Nobuyuki Otsu, 1979. A Threshold Selection Method from Gray-Level Histograms. IEEE Trans. SYSTREMS, MAN, Cybern.

Penuelas, J., Baret, F., & Filella, I, 1995. Semi-empirical indices to assess carotenoids/chlorophyll a ratio from leaf spectral reflectanc. Photosynthetica. 31, 221–230.

Prasad, A.K., Chai, L., Singh, R.P., Kafatos, M., 2006. Crop yield estimation model for Iowa using remote sensing and surface parameters. Int. J. Appl. Earth Obs. Geoinf. 8, 26-33. https://doi.org/10.1016/j.jag.2005.06.002

Qi, J., Chehbouni, A., Huete, A.R., Kerr, Y.H., Sorooshian, S., 1994. A modified soil adjusted vegetation index. Remote Sens. Environ. 48, 119-126. https://doi.org/10.1016/0034-4257(94)90134-1

Raun, W.R., Solie, J.B., Johnson, G.V., Stone, M.L., Lukina, E.V., Thomason, W.E., Schepers, J.S., 2001. In-Season Prediction of Potential Grain Yield in Winter Wheat Using Canopy Reflectance. Agron J. 93, 131. https://doi.org/10.2134/agronj2001.931131x

Reynolds, C.A., Yitayew, M., Slack, D.C., Hutchinson, C.F., Huete, A., Petersen, M.S., 2000. Estimating crop yields and production by integrating the FAO Crop Specific Water Balance model with real-time satellite data and ground-based ancillary data. Int. J. Remote Sens. 21, 3487-3508. https://doi.org/10.1080/014311600750037516

Rondeaux, G., Steven, M., Baret, F., 1996. Optimization of soil-adjusted vegetation indices. Remote Sens. Environ. 55, 95-107. https://doi.org/10.1016/0034-4257(95)00186-7

Rouse JW, Haas RH, Schell JA, Deering DW, Harlan JC, 1974. Monitoring the vernal advancements and retrogradation. Texas, Texas A & M University.

Sankaran, S., Khot, L.R., Espinoza, C.Z., Jarolmasjed, S., Sathuvalli, V.R., Vandemark, G.J., Miklas, P.N., Carter, A.H., Pumphrey, M.O., Richard Knowles, N., Pavek, M.J., 2015. Low-altitude, high-resolution aerial imaging systems for row and field crop phenotyping: A review. Eur J Agron. 70, 112-123. https://doi.org/10.1016/j.eja.2015.07.004

Shanahan, J.F., Schepers, J.S., Francis, D.D., Varvel, G.E., Wilhelm, W.W., Tringe, J.M., Schlemmer, M.R., Major, D.J., 2001. Use of Remote-Sensing Imagery to Estimate Corn Grain Yield. Agron J. 93, 583. https://doi.org/10.2134/agronj2001.933583x

Silleos, N.G., Alexandridis, T.K., Gitas, I.Z., Perakis, K., 2006. Vegetation Indices: Advances Made in Biomass Estimation and Vegetation Monitoring in the Last 30 Years. Geocarto Int. 21, 21-28. https://doi.org/10.1080/10106040608542399

Son, H., Kim, Changmin, Kim, Changwan, Kang, Y., 2015. Prediction of Government-Owned Building Energy Consumption Based on an RReliefF and Support Vector Machine Model. J. Civ. Eng. Manag. 21, 748–760. https://doi.org/10.3846/13923730.2014.893908

Swain, K.C., Thomson, S.J., Jayasuriya, H.P.W., 2010. Adoption of an Unmanned Helicopter for Low-Altitude Remote Sensing to Estimate Yield and Total Biomass of a Rice Crop. Trans. ASABE. 53, 21-27. https://doi.org/10.13031/2013.29493

Thenkabail, P.S., Lyon, J.G., Huete, A., 2016. Hyperspectral Remote Sensing of Vegetation. CRC Press.

Tilly, N., Aasen, H., Bareth, G., 2015. Fusion of Plant Height and Vegetation Indices for the Estimation of Barley Biomass. Remote Sens. 7, 11449-11480. https://doi.org/10.3390/rs70911449

Uno, Y., Prasher, S.O., Lacroix, R., Goel, P.K., Karimi, Y., Viau, A., Patel, R.M., 2005. Artificial neural networks to predict corn yield from Compact Airborne Spectrographic Imager data. Comput. Electron. Agric. 47, 149-161. https://doi.org/10.1016/j.compag.2004.11.014

Valente, M., Prades, A., Laux, D., 2013. Potential use of physical measurements including ultrasound for a better mango fruit quality characterization. J. Food Eng. 116, 57-64. https://doi.org/10.1016/j.jfoodeng.2012.11.022

Vergara-Díaz, O., Zaman-Allah, M.A., Masuka, B., Hornero, A., Zarco-Tejada, P., Prasanna, B.M., Cairns, J.E., Araus, J.L., 2016. A Novel Remote Sensing Approach for Prediction of Maize Yield Under Different Conditions of Nitrogen Fertilization. Front. Plant Sci. 7, 666.

Wang, L. ’ai, Zhou, X., Zhu, X., Dong, Z., Guo, W., 2016. Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. Crop J. 4, 212-219. https://doi.org/10.1016/j.cj.2016.01.008

Yang, H., 2011. Remote Sensing Technique for Predicting Harvest Time of Tomatoes. Procedia Environ Sci. 10, 666-671. https://doi.org/10.1016/j.proenv.2011.09.107

Ye, X., Sakai, K., Manago, M., Asada, S.-I., Sasao, A., 2007. Prediction of citrus yield from airborne hyperspectral imagery. Precis. Agric. 8, 111-125. https://doi.org/10.1007/s11119-007-9032-2

Yue, J., Yang, G., Li, C., Li, Z., Wang, Y., Feng, H., Xu, B., 2017. Estimation of Winter Wheat Above-Ground Biomass Using Unmanned Aerial Vehicle-Based Snapshot Hyperspectral Sensor and Crop Height Improved Models. Remote Sens. 9, 708. https://doi.org/10.3390/rs9070708

Zarco-Tejada, P.J., Miller, J.R., Morales, A., Berjón, A., Agüera, J., 2004. Hyperspectral indices and model simulation for chlorophyll estimation in open-canopy tree crops. Remote Sens. Environ. 90, 463-476. https://doi.org/10.1016/j.rse.2004.01.017

Zhao, D., Raja Reddy, K., Gopal Kakani, V., Read, J.J., Koti, S., 2007. Canopy reflectance in cotton for growth assessment and lint yield prediction. Eur J Agron. 26, 335-344. https://doi.org/10.1016/j.eja.2006.12.001