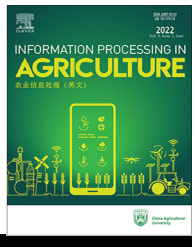
Available at www.sciencedirect.com

INFORMATION PROCESSING IN AGRICULTURE 10 (2023) 361–376

journal homepage: www.keaipublishing.com/en/journals/information-processing-in-agriculture/



An investigation on the best-fit models for sugarcane biomass estimation by linear mixed-effect modelling on unmanned aerial vehicle-based multispectral images: A case study of Australia

Sharareh Akbarian [a,*], *Chengyuan Xu* [b], *Weijin Wang* [c], *Stephen Ginns* [d], *Samsung Lim* [a]

[a] *School of Civil and Environmental Engineering, University of New South Wales, Sydney 2052, Australia*
[b] *School of Health, Medical and Applied Sciences, Central Queensland University, Bundaberg 4670, Australia*
[c] *Science & Technology Division, Department of Environment and Science, Queensland 4017, Australia*
[d] *Queensland Department of Agriculture and Fisheries, Brisbane 4000, Australia*

ARTICLE INFO

Article history:
Received 13 April 2021
Received in revised form
7 March 2022
Accepted 29 March 2022
Available online 1 April 2022

Keywords:
Sugarcane biomass estimation
Unmanned Aerial Vehicle (UAV)
Random effects
Nitrogen fertilizer treatment
Model selection
Vegetation Index (VI)

ABSTRACT

Due to the worldwide population growth and the increasing needs for sugar-based products, accurate estimation of sugarcane biomass is critical to the precise monitoring of sugarcane growth. This research aims to find the imperative predictors correspond to the random and fixed effects to improve the accuracy of wet and dry sugarcane biomass estimations by integrating ground data and multi-temporal images from Unmanned Aerial Vehicles (UAVs). The multispectral images and biomass measurements were obtained at different sugarcane growth stages from 12 plots with three nitrogen fertilizer treatments. Individual spectral bands and different combinations of the plots, growth stages, and nitrogen fertilizer treatments were investigated to address the issue of selecting the correct fixed and random effects for the modelling. A model selection strategy was applied to obtain the optimum fixed effects and their proportional contribution. The results showed that utilizing Green, Blue, and Near Infrared spectral bands on models rather than all bands improved model performance for wet and dry biomass estimates. Additionally, the combination of plots and growth stages outperformed all the candidates of random effects. The proposed model outperformed the Multiple Linear Regression (MLR), Generalized Linear Model (GLM), and Generalized Additive Model (GAM) for wet and dry sugarcane biomass, with coefficients of determination (R^2) of 0.93 and 0.97, and Root Mean Square Error (RMSE) of 12.78 and 2.57 t/ha, respectively. This study indicates that the proposed model can accurately estimate sugarcane biomasses without relying on nitrogen fertilizers or the saturation/senescence problem of Vegetation Indices (VIs) in mature growth stages.

* Corresponding author at: School of Civil and Environmental Engineering, University of New South Wales, Oval Ln, Kingsford, NSW 2052, Australia.
E-mail address: s.akbarian@unsw.edu.au (S. Akbarian).
Peer review under responsibility of China Agricultural University.
https://doi.org/10.1016/j.inpa.2022.03.005

1. Introduction

Nowadays, confronting the demands for agricultural products as the world population increases is a major challenge. Sugarcane, a perennial crop suitable to grow in tropical and subtropical areas contributes to nearly 80% of the global sugar-based products [1]. Farmers together with food industries are seeking to monitor sugarcane growth in an accurate and efficient way to ensure the quality and quantity of yields before harvesting [2]. Biomass is crucial for the monitoring of crop growth and yield prediction [3]. Accurate estimation of biomass through a precise and non-destructive way requires data on the crop status with high spatial and temporal resolution [4].

The recent advancements of remote sensing techniques have proved the capability in the estimation of sugarcane biomass [5]. Despite the widespread use of satellite-based data for large spatial coverage [6-9], there are some concerns about the low spatial–temporal resolution, the high cost, and the weather constraints of monitoring biomass at the regional scale [10]. The Unmanned Aerial Vehicle (UAV) platform has been broadly employed to estimate sugarcane biomass because of its advantages, such as low operation cost, high spatial resolution (<10 cm), and high temporal frequency [2,11–14]. Although these models have proven to be reliable in predicting sugarcane biomasses at the field or block level, they have not been investigated at the finer scales. This is a critical issue that cannot be ignored since sugarcane production varies not only seasonally and geographically but also within crop rows.

In addition to a remote sensing platform, various regression methods have also been implemented to stablish crop biomass estimation models. Multiple Linear Regression (MLR) is the most widely used conventional statistical regression method for determining a fixed linear relationship between remote sensing data and biomass contents [15,16]. Furthermore, machine learning methods such as Artificial Neural Network (ANN), Support Vector Machine (SVM), and Random Forest Regression (RFR) models are being employed more frequently to improve the nonlinear estimation capability of the crop biomass models [15,17–20]. However, most of these methods have attempted to define the relationships between the crop variables as predictor variables (e.g., spectral information) and response variables (e.g., biomass) [21]. The specific geographical and environmental factors that affect canopy reflectance, such as plant species, growth stage, and location, could not be taken into account by these methods [22,23]. The performance of the existing biomass estimation models varies depending on random factors (e.g., crop growth stage, fertilization level, site location, and plant species) [21], since the fixed-effect regression models do not consider these factors [24]. The mixed-effect regression models, which may generalize correlation models to various crop types and growth phases, can solve this challenge for a wide range of applications [25–27].

One prominent challenge in the application of the mixed-effect models is the choice of the appropriate random effects. The random effects are usually assigned to grouping factors as a sample of all possibilities for which we tend to generalize results to a whole population based on the representative sampling [28]. This grouping can directly affect the performance of the mixed-effect models in highly structured data with correlated observations and repeated measurements taken on the same experimental unit [29,30]. The analysis of repeatedly acquired agricultural data from the same sample units at different time periods, referred to as longitudinal data, could demonstrate how the sample units evolve over time. The mixed-effect models can quantify the relationships between independent and dependent variables in longitudinal studies with repeated measurements, by considering both fixed and random effects appropriately [27]. However, to the best of our knowledge, no previous study has inspected the effect of various random effects on biomass estimation in a sugarcane field.

In addition to selecting a suitable regression method and its parameters, the finding of a proper subset of explanatory variables is an important step toward establishing and determining the accuracy of the biomass estimation model. There are many explanatory variables retrieved from remote sensing and ground data, such as the reflectance of spectral bands, textural information, and vegetation indices (VIs), which can be used for biomass estimation [15]. VIs such as the Normalized Difference Vegetation Index (NDVI) [31], have been repeatedly used to measure biomass [1,32], and provided reliable information about crop growing status [33,34]. Sankaran et al. [35] estimated the dry biomass of beans by using the Green Normalized Difference Vegetation Index (GNDVI) and reported a high correlation between the mean GNDVI and the biomass. Rahman and Robson [36] also used the GNDVI extracted from multi-temporal satellite data to predict sugarcane crop yield in Bundaberg, Australia. Hunt et al. [37] presented a good correlation between biomass contents and GNDVI using UAV-based multispectral imagery. The Normalized Difference Red Edge (NDRE) as a variant of the NDVI calculated by substituting Red band with Red Edge (RE) band, has been identified as the sugarcane biomass predictor with the highest sensitivity [38].

Contrary to the broad applicability of VIs, the saturation problem known as the imbalance between the Red and Near-infrared (NIR) reflection that occurs because Red reflectance becomes flat when the leaf area index is high, may lead to inaccurate biomass estimation especially at the peak stage of crop growth and on large biomass crops, such as sugarcane or mature crops [39]. Leaf senescence also makes the use of VIs difficult because the photosynthetic capacity of leaves declines with time during the late cropping season [40,41] and, therefore, the correlation between a VI and biomass is low. To overcome this issue, we focus directly on the reflectance measurements in five spectral bands, namely Red, Green, Blue, NIR, and RE from a multi-temporal of UAV-based multispectral images, rather than using a particular VI.

Selection of the most efficient spectral bands as explanatory variables is important in order to remove any irrelevant variables, reduce the dimension of variable space, and

increase model performance [42]. Stepwise regression methods have been widely used in many studies to model biomass yield [43], in order to simplify models with the minimum number of explanatory variables. However, they are not appropriate when the number of variables are excessive, or observations are lack [44]. The variable selection methods based on the Akaike Information Criterion (AIC) have been used to overcome the weaknesses of the stepwise selection methods [42]. They can be applied to the mixed-effect models, linear and non-linear models, and normally or non-normally distributed data [45].

In this research, similar to the work of Wang et al. [21], we attempted to develop a mixed-effect model based on the ground measurements coupled with the remote sensing observations. The multi-temporal UAV images with a high spatial resolution (<3 cm) were assessed in this study to develop a more accurate biomass estimation at the row level. Additionally, we tested the model's performance for different random effects and extended to the model selection strategy to use prominent spectral bands in order to improve the accuracy of biomass estimation in the sugarcane field. We used the ground measurements of the wet and dry biomasses and the UAV images, at five separate months from different stages of sugarcane growth, from 12 plots treated with three different nitrogen fertilizers, over the complete 2018–2019 growth cycle. Because the plant's water content is directly influenced by soil moisture and the weather, dry biomass is also considered as a more reliable measure than wet biomass in yield prediction [46]. Nitrogen fertilizer treatments play a key role in the biomass contents, and the crop growth consequently. Hence, the availability of information regarding the applied nitrogen fertilizer treatments could influence the accuracy of the biomass estimation. Thus, we investigated: (1) the significance of random factors to the sugarcane biomass estimation and the performance of the proposed model. This would contribute to bridging the knowledge gap in the use of the appropriate random factor in a longitudinal sugarcane biomass estimation study that utilizes repeating measurements; (2) the possibility to estimate the biomass without any knowledge of the applied nitrogen fertilizer to the crop; (3) the efficiency of the extracted spectral information in estimating the sugarcane biomass; and (4) the most appropriate explanatory variables from the spectral information to address the VIs' saturation problem at the peak growth and the leaf senescence at the harvest stage.

2. Materials and methods

2.1. Study area

Sugarcane grows best in well-drained/fertilized soil with around 1 500 mm of annual rainfall or irrigation in a sunny and warm climate. These ideal conditions can be found along a 2 000 km strip of land on the east coast of Australia, with about one-third of the sugarcane farms located in the northern Queensland. This research was carried out at an existing experimental site where different nitrogen fertilizer management practices were trialed for sugarcane farming in the Bundaberg region, Queensland (152° 24′ E, 24° 50′ S, see Fig. 1). Bundaberg has a subtropical humid climate with hot, rainy summers and mild, dry winters, and accounts for about 28 percent of the total sugar output in Australia [47]. The average yearly precipitation is 1 143 mm which is mostly from October to March. The average daily temperature ranges from 9.9 °C to 30.3 °C.

Sugarcanes of the Q242 variety were planted in November 2015. The nitrogen fertilizer experiment was established in December 2016, with 12 treatments arranged in 48 plots in a randomized block design with 4 replicates [48]. Each plot is approximately 21 m long and 11 m wide, with six crop rows spaced by 1.83 m between rows and plots.

2.2. Field measurements

Sugarcane samples were collected from 12 out of the 48 plots (Fig. 1) to measure the dynamics of plant biomass and nitrogen uptake in 3 different treatments (for details see Table 1) during the 2018–2019 crop growing season [48]. Wet and dry biomass samples were collected at different stages of sugarcane growth to observe the temporal variation from tillering to mature crop (Table 2). Since this is the third ratoon crop, there is no germination and establishment phase; only tillering, ground growth, and ripening phases are present. These data serve as a source of ground truth for adjusting and validating biomass estimating models. The UAV imagery was then acquired for these plots at dates close to the plant sampling events. The time intervals (5–8 days except the first sampling) between the two data collection methods are short in comparison to the nine-month span of all data collection (January to September), hence such a discrepancy is assumed to not affect the conclusions of the study.

As each plot had six rows, Rows 2 to 5 were assumed to be equally representative of the crops in the same plot, whilst Rows 1 and 6 may be affected by fertilizer contamination of the neighboring plots and, thus, were excluded from the analysis. Rows 2 and 5 were sampled for the ground measurements in January, February, April and June 2019, and Rows 3 and 4 remained intact until the final harvest date. As these central two rows had not been affected by the in-season sampling, they were assumed as the best representatives of the crops at the harvest season (September 2019).

The plant samples were taken on six occasions from January to September 2019 throughout the cropping season. From January to June 2019, the aboveground plant samples were taken from a 1 m section in Rows 2 and 5 of each plot. The number of plants and the total biomass weight in each section were recorded. In September 2019, the crops were harvested from two 5 m sections in the middle two rows to weight the wet biomass. Immediately after the total biomass weighing during the harvest, about 1–20 stalks (usually 10 each from each row) were randomly selected and segregated into stalks and cabbage/leaves. The stalks and cabbage/leaves were weighed separately, and then crushed and cut into small pieces using a mulcher, subsampled, and dried at 60 °C for more than 48 h for determination of water contents. Wet biomass yield in each plot was calculated as the sum of fresh stalk and leaf/cabbage, corrected for the crop density (number of stalks per meter measured over a 10 m section in Row 3 or

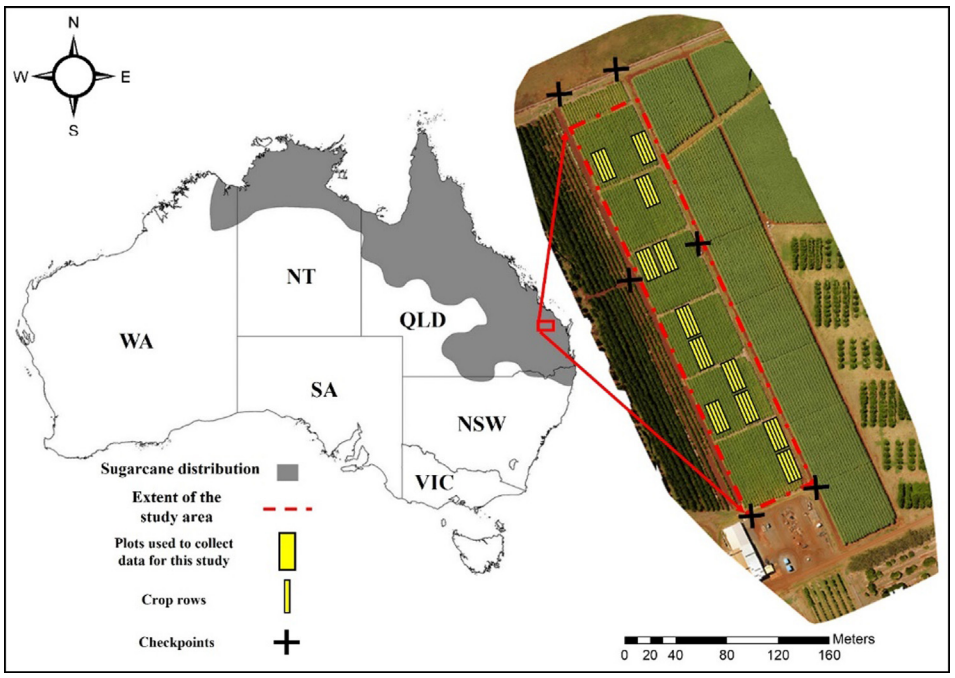


Fig. 1 – Sugarcane field distribution in Australia (left) and the study area (right).

Table 1 – Specifications of the nitrogen fertilizer treatments.

Nitrogen fertilizer treatment	Fertilizer ratio	Nitrogen rate/(Kg/ha)
Treatment 1	Control	0
Treatment 3	Urea	130
Treatment 9	75% Polymer-coated urea + 25% urea	130

Table 2 – Dates of data collection.

Survey number	Plant sampling date	UAV-survey date	Sugarcane growth phase
1	03 January 2019	31 January 2019	Tillering
2	04 February 2019	10 February 2019	Tillering
3	08 April 2019	31 March 2019	Ground growth
4	18 June 2019	11 June 2019	Ground growth
5	24 September 2019	29 September 2019	Ripening

4). The dry biomass yield was estimated after correction for water content.

Fig. 1 shows the polygons of the selected plots, which were generated in ArcMap (version 10.7.1) to extract the average numeric values of the pixels that fell inside each crop row. The zonal statistics function of ArcToolbox was used to generate a mean of the reflectance values in the Red, Green, Blue, NIR, and RE from the UAV images. The mean values were saved in a 2D dataset table along with the wet and dry biomasses, which were also measured at the same polygon areas. Figs. 2 and 3, and Table 3 display the descriptive statistics on wet and dry biomasses, as well as the five spectral bands throughout the whole season.

2.3. *UAV imagery*

The multi-temporal collection of multispectral images with a spatial resolution of 3 cm was acquired at different stages of sugarcane growth, using a UAV-integrated MicaSense RedEdge sensor (Micasense, Seattle, WA, USA) on a Phantom 3 Advanced multirotor drone (DJI, Shenzhen, China). The images were captured at a height of 45 m above ground and with a parallel camera CCD angle to the ground. The 60% and 90% of side and forward overlap in all flights respectively, generated a satisfactory performance of image stitching. A downwelling light sensor and a Global Positioning System (GPS) antenna integrated with the sensor were used to collect

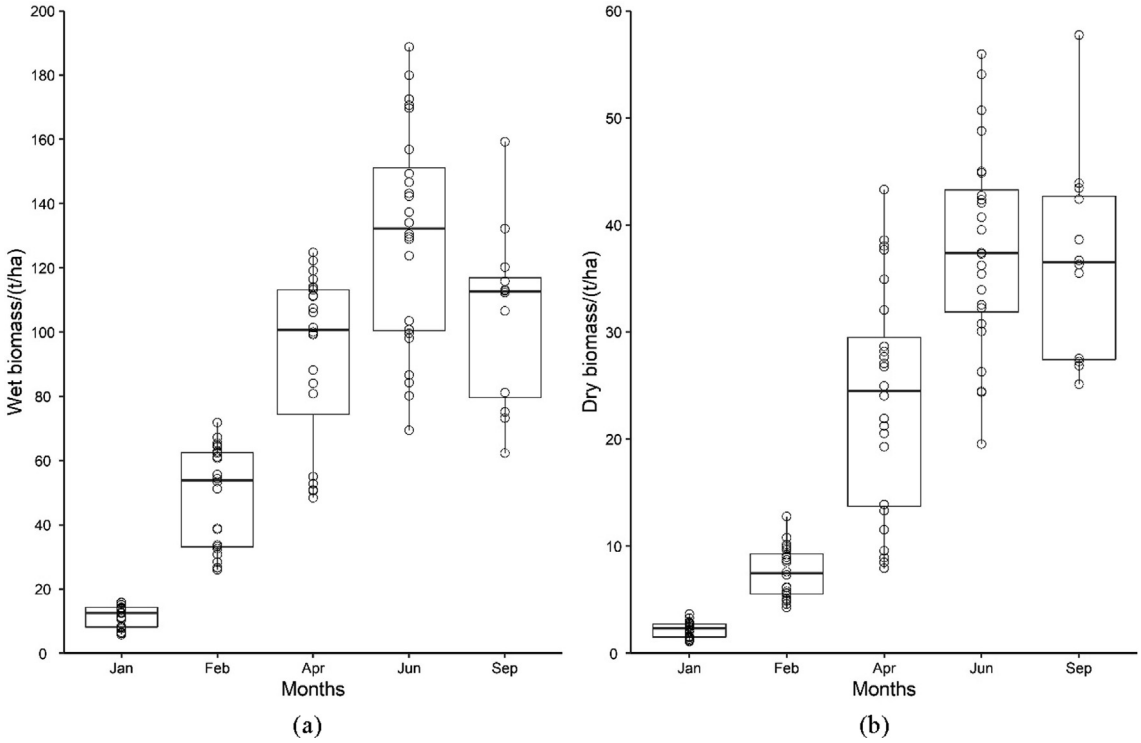


Fig. 2 – The summary statistics for (a) wet and (b) dry biomasses over the complete period of data collection.

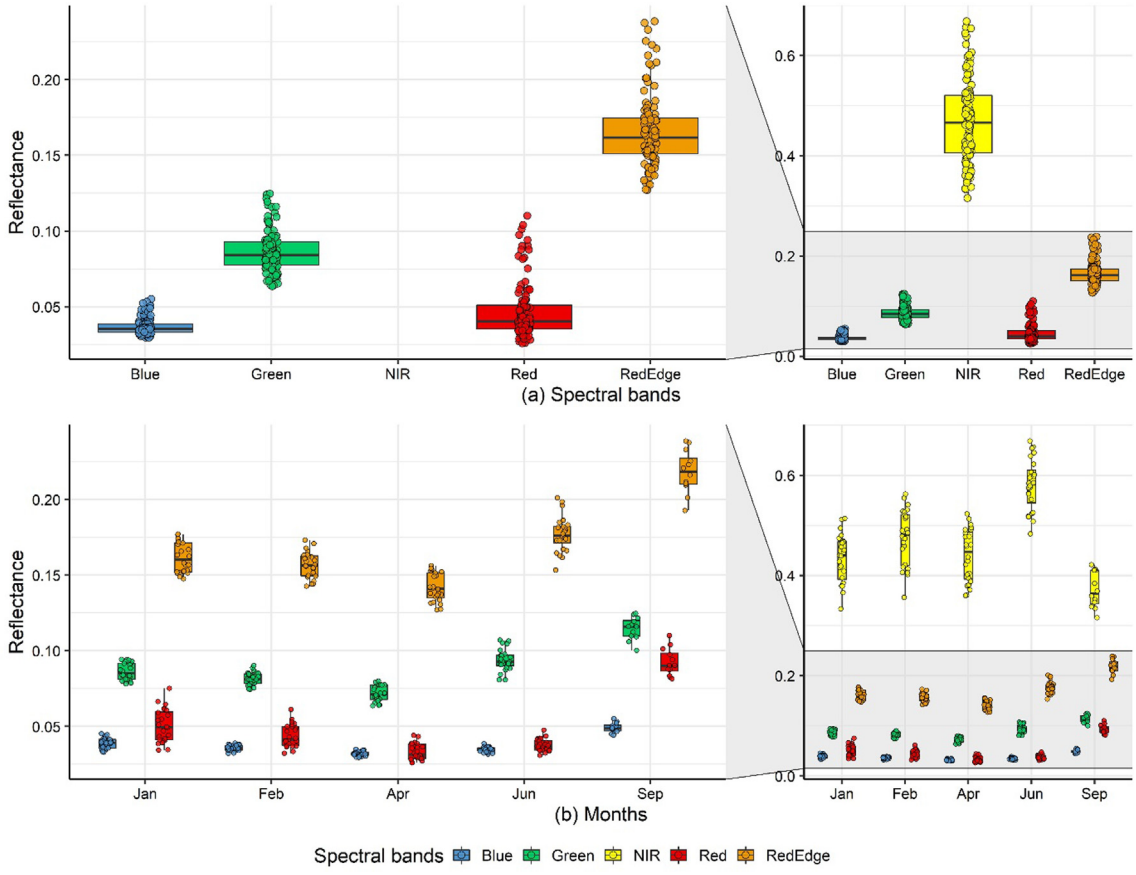


Fig. 3 – The summary statistics for the five spectral bands (a) over the entire period of data collection and (b) per month with a zoom-in to the further details.

Table 3 – Descriptive statistics of wet and dry biomasses and spectral bands across the entire season.

	Minimum	Mean	Maximum	Median
Wet biomass/(t/ha)	5.81	74.92	188.74	70.62
Dry biomass/(t/ha)	1.105	20.136	57.734	16.597
Red	0.026	0.047	0.11	0.040
Green	0.063	0.086	0.124	0.084
Blue	0.029	0.036	0.055	0.035
NIR	0.315	0.469	0.522	0.465
RE	0.127	0.165	0.238	0.161

irradiance and location data while UAV images were collected between 11 am and 2 pm (−1 to +2.5 hrs. from solar noon) to minimize any shadowing impact. Afterward, The UAV spectral reflectance was calibrated using the information of irradiance, heading, orientation, and solar angle. The image acquisition dates are presented in Table 2. The multispectral images include reflectance measurements in five spectral bands: Blue (wavelength of 460–510 nm), Green (545–575 nm), Red (630–690 nm), RE (712–722 nm), and NIR (820–860 nm). Six ground checkpoints were built and marked with a Real-time Kinematic (RTK) GPS (Leica Geosystems CS20 controller plus GS14 antenna, Hexagon, Madison, AL, USA) across the field study (see Fig. 1). Later, the ground checkpoints were utilized to geo-reference each sample data in the World Geodetic System (WGS) 1984 projection, Universal Transverse Mercator (UTM) Zone 55 datum, and then, to register the ground data with the corresponding multispectral images.

The acquired images were orthorectified and mosaicked on Pix4Dmapper software package (Pix4D S. A. Prilly, Switzerland), designed to process the multispectral images using computer vision and photogrammetry methods. The GCPs were used to geo-reference the study area, increase the accuracy, and decrease the misalignment on the Pix4D software.

2.4. Analysis of the vegetation indices

An overall analysis of the most frequently used VIs in biomass estimation was conducted to understand the VIs' performance, concerning the saturation problem at the peak growth and the senescence problem at the harvest. For this, the NDVI, NDRE, and GNDVI were derived from the UAV imagery corresponding to each of the growth stages using Eq. (1), Eq. (2), and Eq. (3). To confirm the limitations of these indices for biomass estimation of a mature crop, the linear, polynomial (second order), logarithmic, and exponential regression models were fitted to scatterplots of the VIs versus wet and dry biomasses. The Coefficient of Determination (R^2) and the Root Mean Square Error (RMSE) were used to analyze the performance of the four models. The statistical analyses were executed on RStudio version 1.3.959 [49].

$$NDVI = \frac{(NIR - Red)}{(NIR + Red)} \tag{1}$$

$$NDRE = \frac{(NIR - Rededge)}{(NIR + Rededge)} \tag{2}$$

$$GNDVI = \frac{(NIR - Green)}{(NIR + Green)} \tag{3}$$

2.5. Mixed-effect models

To evaluate the capability of the spectral information to estimate wet and dry sugarcane biomasses, we investigated the relationship between the response and predictor variables using the mixed-effect models. It is a regression model where some of the predictor variables are random, and others are fixed. The matrix form of the models is given in Eq. (4), where y represents a column vector standing for the response variable, X and Z represent design matrices of the predictor variables, a is an unknown vector of the fixed-effect regression coefficients, b represents a vector of the random effects, and ε represents a vector of the residuals [50].

$$y = aX + bZ + \varepsilon \tag{4}$$

However, different choices of the random effects influence the performance of the biomass estimation models. For example, using the minimum level of random effects decreases data variance due to the smaller number of group combinations and influences the model's robustness [21]. For both wet and dry sugarcane biomass estimations, the five spectral bands were considered as the predictor variables with the fixed effects and the wet and dry biomasses as the response variables. The contributions of random factors were investigated by using different selections of random effects (as explained in the following Section).

The Analysis of Variance (ANOVA) function is used to test for the predictor of the fixed-effect variables on the total abundance of the response variable, taking the optimal random effects into account. First, the full model was fitted by the maximum likelihood and a second model lacking the fixed effects of the explanatory variables. Then, these two models for both wet and dry biomass estimations were compared with a likelihood ratio test using the ANOVA function.

2.5.1. Random effects

To analyze the performance of the mixed-effect model in response to different selections of the random effects, 108 observations from the 12 plots were considered for the estimations of wet and dry biomasses. These were repeated observations at different growth stages summing up a total of 24 samples in January, 24 samples in February, 24 samples

in March and April, 24 samples in June (all from Rows 2 and 5 as described in Section Field measurements), and 12 samples in September from Rows 3 and 4 (see Section 2.2). Thus, the plots and the growth stages of the samples are treated as random effects with 12 and five levels, respectively. We randomly selected three of the 12 treatments applied to the experimental field to generalize their contributions to the sugarcane's wet and dry biomasses. In both wet and dry biomass estimation models, the predictor variables represent the spectral bands, the response variable represents the wet and dry biomasses, set as fixed effects. The random effects include different combinations of the plots (ID-12 levels), the data collection dates from different sugarcane growth stages (M-five levels), and the nitrogen fertilizer treatments (T-three levels). The R^2 and RMSE were used to compare the predictive capabilities of the estimation models. The statistical analyses were executed on RStudio version 1.3.959 [49].

2.6. Analyzing different models

In our analysis, one linear regression model called the Multiple Linear Regression (MLR), and two data-driven methods, namely the Generalized Linear Model (GLM) and the Generalized Additive Model (GAM), were used to construct prediction models and analyze our mixed-effect model. The MLR is a mathematical algorithm that uses multiple predictor variables to obtain the response value. Along with their generality, MLR models have been widely used in many biomasses' estimation and yield prediction studies [51–54]. The GLM is an extended form of the conventional linear model that allows the specification of models whose response variable follows different distributions. We chose poisson regression among various distributions (e.g., the normal, binomial, poisson, and gamma) because of its advantage in fitting the logarithmic model of variables [55]. The GAM developed by Chambers et al. [56] is essentially an extension of the GLM, but it is not restricted by linear relationships. It is particularly useful in revealing the non-linear effects of the environment variables on the biomass [57]. In this research, the "mgcv" package of R is used [58]. The R^2 and RMSE between fitted and observed data were used as accuracy evaluation indices of these models with the proposed mixed-effect model here. The statistical analyses were performed on RStudio version 1.3.959 [49].

2.7. Model selection

A proper model selection algorithm is required to identify the most appropriate explanatory variables from the spectral information and overcome the saturation and senescence problem of mature growth stages. In this research, the "Dredge" function of an R package "MuMIn" [59] was used to select the most proper variables for the model, as the fixed-effects terms. The selected variables were used as the final wet and dry sugarcane biomass estimation models. This analysis generated a selection table of models with subsets of fixed-effects terms in the model by adding and removing the predictor variables in each step and evaluated the prediction performance with respect to the AIC. This process was repeated until all subsets were tested by running all combinations of the variables examined. The number of possible subsets is n^2, where n is the number of explanatory variables. All data processing and statistical analyses were performed using RStudio version 1.3.959 [49].

2.8. Evaluation of wet and dry biomass models

To evaluate the performance of the final estimation models of wet and dry biomasses, the correlation test between the predicted and actual wet and dry biomasses was performed using the statistical parameters. Adjusted R^2 (AR^2) and RMSE were used to evaluate the performance of the final models. Generally, higher AR^2 and lower RMSE values indicate a better performance of the estimation model.

3. Results and discussion

3.1. Results

3.1.1. Analysis of Vegetation indices

Fig. 4 displays the overall analysis of three VIs versus wet and dry biomasses, as well as the effect of fertilizer treatments on biomass values. All three VIs perform similarly in the early to mid-season which includes January, February, March, and April, and the variance in crop response to the different treatments is less noticeable than in the late season which includes June and September. However, by the mid-season, a substantial difference in crop response to each treatment could well be noticed. Treatment 9 (the triangle symbol in Fig. 4) which includes 75% Polymer-coated urea, 25% urea and 130 kg/ha Nitrogen, has produced the highest amount of biomass in June and September when compared to other treatments. In the late season, there was a low correlation between the NDVI values and the wet and dry biomasses in September due to the leaf senescence, and equally low correlation between the three nitrogen treatments compared to the earliest growth stages (Fig. 4(a) and Fig. 4(b)). The NDRE and GNDVI exhibited higher resistance to senescence than NDVI in conditions of higher biomass (Fig. 4(c) to Fig. 4(f)). The included spectral bands in each VI might account for this performance difference. The Red and NIR bands, as discussed in earlier sections, may cause biomass estimation to be incorrect at the peak growth stage. Therefore, the mathematical combination of these bands in the NDVI resulted in a lower correlation between the VI and biomass as compared to the NDRE and GNDVI, which lack the Red band in their formulae. However, the senescence and saturation problems were still present in the samples from June and September measurements, when the crop reached the highest amount of biomass.

Moreover, the polynomial model resulted in the highest R^2 and the lowest RMSE, outperforming the other regression models (Table 4 and Table 5). A polynomial model regression of the observed wet and dry biomasses with the NDVI produced the models with R^2 of 0.41 and 0.40, respectively. With the NDRE values, it yielded R^2 of 0.32 and 0.31, respectively. And with the GNDVI values, the R^2 values were 0.32 and 0.30, respectively.

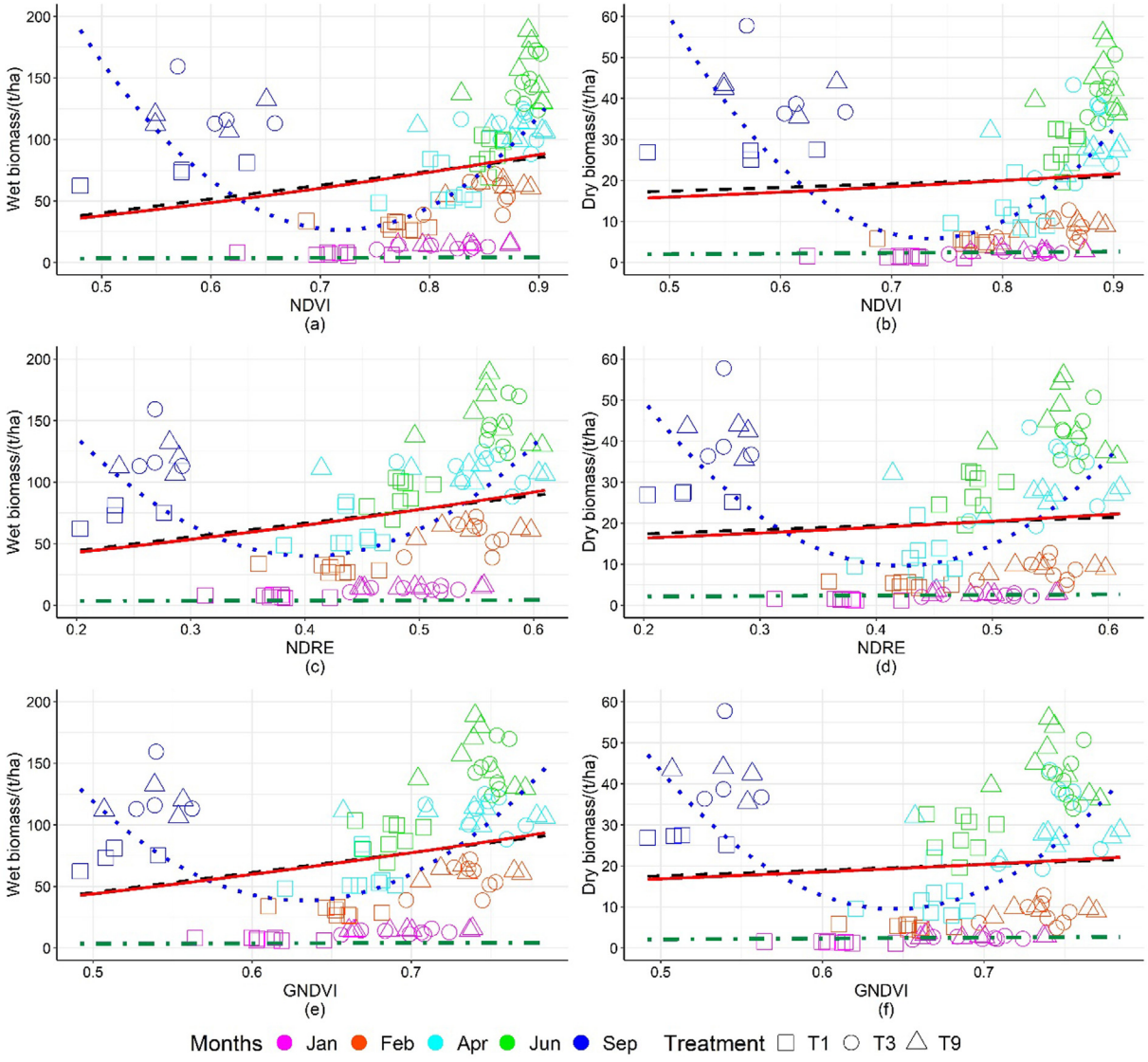


Fig. 4 – The scatterplots of (a) wet biomass vs NDVI (b) dry biomass vs NDVI (c) wet biomass vs NDRE (d) dry biomass vs NDRE (e) wet biomass vs GNDVI (f) dry biomass vs GNDVI.

Table 4 – Performance of regression models for estimating wet biomass.

VI	Model	Regression equation	R^2	RMSE/(t/ha)
NDVI	Exponential	$y = -63.25e^{61.38\ x}$	0.06	48.08
	Linear	$y = 112.63\ x - 15.96$	0.05	48.47
	Logarithm	$Y = 65.16 \ln x + 89.44$	0.02	48.92
	Polynomial	$y = 307.81\ x^2 + 112.44\ x + 74.92$	0.41	38.15
NDRE	Exponential	$y = -56.26\ e^{81.29\ x}$	0.06	48.07
	Linear	$y = 111.67\ x + 22.03$	0.05	48.41
	Logarithm	$y = 30.78 \ln x + 98.77$	0.03	49.09
	Polynomial	$y = 263.11\ x^{-2} + 115.35\ x + 74.91$	0.32	41.11
GNDVI	Exponential	$y = -106.87\ e^{91.48\ x}$	0.06	48.07
	Linear	$y = 161.76\ x - 35.77$	0.05	48.31
	Logarithm	$y = 89.50 \ln x + 109.41$	0.04	48.67
	Polynomial	$y = 265.73\ x^2 + 119.25\ x + 74.91$	0.32	40.84

Table 5 – Performance of regression models for estimating dry biomass.

VI	Model	Regression equation	R^2	RMSE/(t/ha)
NDVI	Exponential	$y = 4.51e\ 6.94^{x}$	0.01	16.23
	Linear	$y = 8.83\ x + 13$	0.05	16.27
	Logarithm	$Y = 0.78 \ln x + 20.31$	0.01	16.29
	Polynomial	$y = 106.2\ x^2 + 8.82\ x + 20.14$	0.40	12.58
NDRE	Exponential	$y = 4.60\ e^{9.62\ x}$	0.01	16.22
	Linear	$y = 9.94\ x + 15.40$	0.01	16.26
	Logarithm	$y = -\ 0.88 \ln x + 19.46$	0.01	16.30
	Polynomial	$y = 92.63\ x^{.2} + 10.32\ x + 20.13$	0.31	13.54
GNDVI	Exponential	$y = 0.92\ e^{9.67\ x}$	0.01	16.24
	Linear	$y = 13.74\ x + 10.74$	0.01	16.26
	Logarithm	$y = 3.92 \ln x + 21.64$	0.01	16.28
	Polynomial	$y = 91.98\ x^2 + 10.12\ x + 20.13$	0.30	13.59

However, the performance of this model proved to be relatively poor in the implementation of the biomass estimation at the peak growth/harvest stage. These results indicate the importance of selecting proper spectral information in biomass estimation to overcome the VIs' high saturation and less correlation with leaf senescence during the medium to high crop biomass conditions. Therefore, we employed five spectral bands individually into the mixed-effect model to enhance the estimations of both wet and dry sugarcane biomasses.

3.1.2. *Mixed-Effect models*

3.1.2.1. Random effects. Identifying the appropriate parameters of the mixed-effect models from which the accurate estimation model can be constructed is the key to success. Eight selections of the random effects were tested, and the optimal models were obtained with respect to the maximum R^2 and minimum RMSE (Table 6). For the wet biomass, the combinations of [ID, IDM] and [ID, IDM, M] resulted in the maximum R^2 (0.98) and the minimum RMSE (5.41 t/ha). For the mixed-effect models on dry biomass, the maximum R^2 (0.97) and the minimum RMSE (2.53 t/ha) were achieved by the combination of [ID, IDM]. Due to the prediction accuracies, the combinations of [ID, IDM] and [ID, IDM, M] with the highest performance were concluded as the optimum random factors of wet and dry biomass estimations on the mixed-effect models.

From the ANOVA results (Table 7), strong evidence of an association was found in all four models by taking the optimal random effects into account for the explanatory variables. Improving the AIC parameter and *p*-value indicates the impact of modification by the intervention of the explanatory variables with fixed effects.

3.1.3. *Analyzing different models*

By comparing the values of R^2 and RMSE, the proposed mixed-effect models for both wet and dry biomasses exhibited significantly better performance compared to the MLR, GLM, and GAM models (Fig. 5 and Fig. 6). The R^2 values for both wet and dry biomass estimations were remarkably similar, with 0.98 and 0.97, respectively. In comparison to the three other techniques, the RMSE values of wet and dry biomass estimations reduced to 5.6 t/ha and 2.62 t/ha, respectively.

Table 6 – Comparison of the predictive capability of the models.

Biomass	Random effects selections	R^2	RMSE/(t/ha)
Wet Biomass	ID	0.60	23.65
	ID, IDM	0.98	5.41
	ID, IDM, M	0.98	5.41
	ID, M	0.92	12.67
	ID, T	0.63	23.52
	ID, IDM, T	failed to converge	
	ID, IDM, M, T	failed to converge	
	ID, M, T	0.92	12.81
Dry Biomass	ID	0.54	7.96
	ID, IDM	0.97	2.53
	ID, IDM, M	0.97	2.62
	ID, M	0.89	4.94
	ID, T	0.60	7.87
	ID, IDM, T	failed to converge	
	ID, IDM, M, T	failed to converge	
	ID, M, T	failed to converge	

Note: ID represents the plots in 12 levels; M represents the data collection dates from different sugarcane growth stages; T represents the nitrogen fertilizer treatments in three levels; IDM represents the combination of the plots and the growth stages.

Table 7 – Results of a likelihood ratio test using the ANOVA function.

Models	AIC	*p*-value
wet biomass ~ random effects (ID, IDM)	994.63	–
wet biomass ~ fixed effects + random effects (ID, IDM)	985.02	0.001**
dry biomass ~ random effects (ID, IDM)	798.78	–
dry biomass ~ fixed effects + random effects (ID, IDM)	775.22	2.907e-06***
wet biomass ~ random effects (ID, IDM, M)	908.44	–
wet biomass ~ fixed effects + random effects (ID, IDM, M)	905.91	0.01*
dry biomass ~ random effects (ID, IDM, M)	709.69	–
dry biomass ~ fixed effects + random effects (ID, IDM, M)	708.45	0.04*

Note: Significance levels are marked as 0 < *p*-value < 0.001 (***), 0.001 < *p*-value < 0.01 (**), and 0.01 < *p*-value < 0.05 (*).

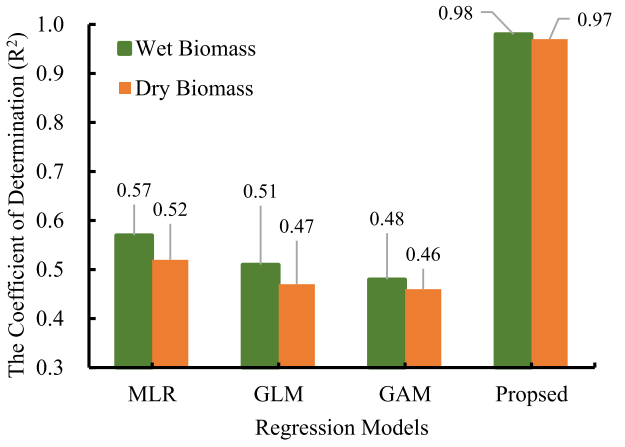


Fig. 5 – The Coefficient of Determination (R^2) of wet and dry biomass for the four regression models.

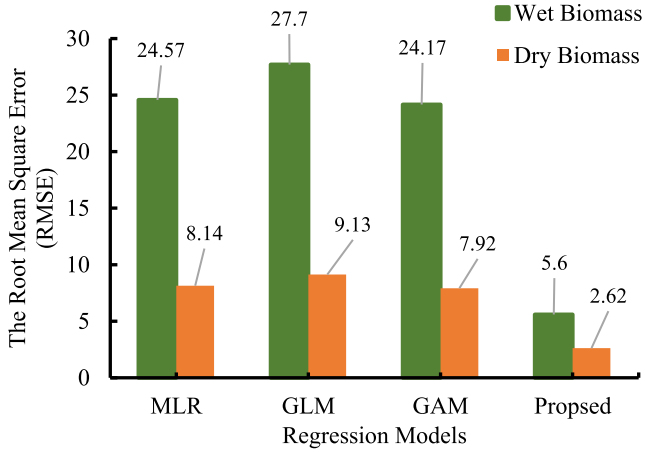


Fig. 6 – The Root Mean Square Error (RMSE) of wet and dry biomass for four regression model.

Both combinations of [ID, IDM] and [ID, IDM, M] produced the similar performance of wet and dry biomass estimations on the model. As a summary, in the associated Figures, we just displayed the outcome of the [ID, IDM, M] combination.

3.1.4. *Model selection*

The proposed mixed-effect models which were constructed based on (ID, IDM) and (ID, IDM, M) as random effects were used to extend the model selection approach to the explanatory variables. When the combination of [ID, IDM] was incorporated as a random effect, the model selection resulted in singularity. To prevent this, the combination of [ID, IDM, M] with equally high predictive performance was chosen to enter the model selection. The AIC values were used as a criterion for choosing the best explanatory subsets among the spectral bands in the proposed models. A lower AIC implies that the model is the most reliable. Tables in Appendix illustrate the selection results for wet and dry biomasses, respectively, starting with the full data set of the spectral bands.

By using the full data set, the result for the wet biomass presented an AIC of 933.0. After analyzing all possible combinations (Appendix Table A1), Red and RE bands were excluded from the best model, and the AIC value improved to 930.8, meaning that Green, Blue, and NIR bands were the optimal spectral value. Eq. (5) shows the final selected spectral bands that are more suited to describe the relationship between spectral information and wet biomass based on the mixed-effect model with ID, IDM, and M as random effects. In this equation B_W stands for the wet biomass, and G, B, and NIR represent the Green and Blue, Near Infrared bands, respectively.

$$B_W = -45.13 - 1301.00G + 4322.50B + 161.79NIR \quad (5)$$

The model selection analysis for the dry biomass was examined with a full set of spectral bands and 32 different combinations of the bands (Appendix Table A2). For the best-selected model with ID, IDM, and M as random effects, when the Red and RE bands were removed the AIC improved from 709.5 at the initial step to 706.1. As for Eq. (3), the final combination of the Green, Blue, and NIR bands will reflect the relationship between the spectral information and the dry biomass. In the formula given in Eq. (6), B_D stands for dry biomass.

$$B_D = -10.52 - 298.21G + 901.85B + 53.16NIR \quad (6)$$

3.1.5. *Evaluation of wet and dry biomass models*

After the model selection, we obtained two estimation models with the selected spectral bands as the fixed effects, and the combination of ID, IDM, and M as the random effects for wet and dry biomasses as shown in the Eq. (5) and Eq. (6). Fig. 7 presents the performances of the two proposed models, expressed by scatterplots between the observed and predicted values of the wet and dry biomasses. The R^2 values and the RMSE show the potential of the obtained models to

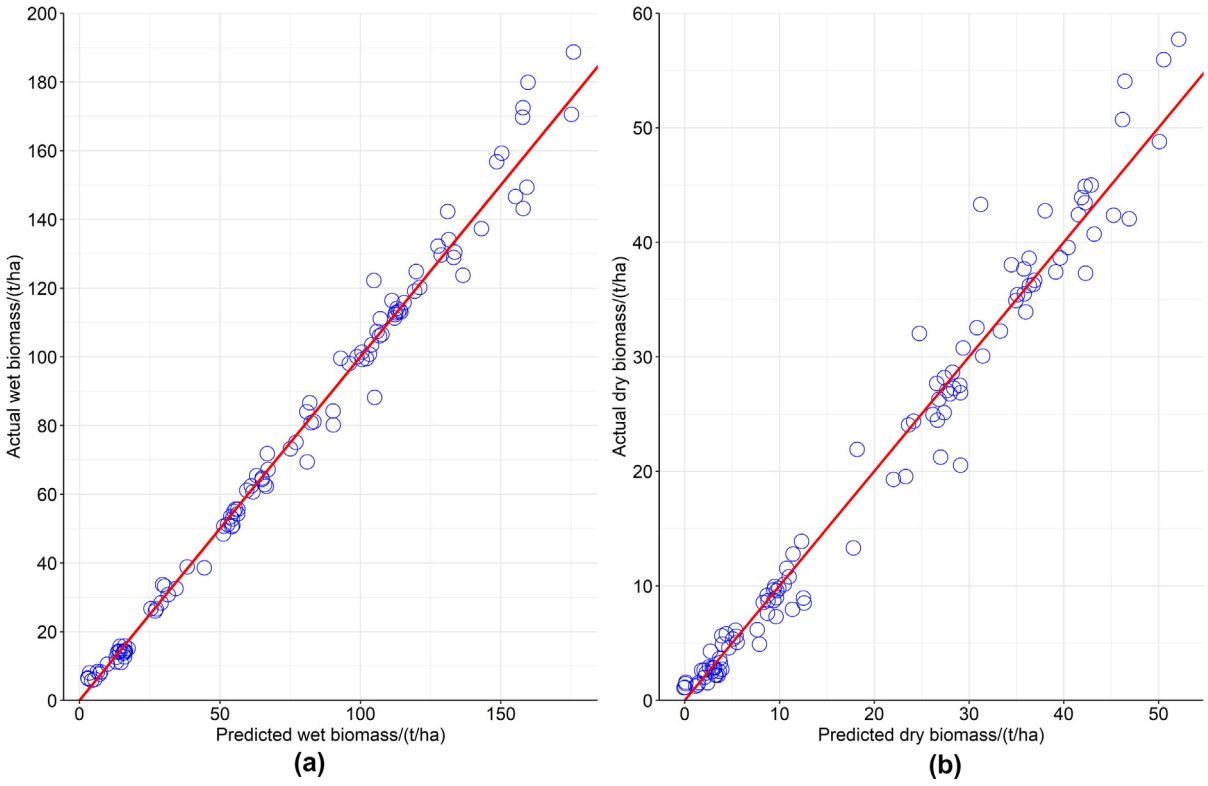


Fig. 7 – Scatterplots of predicted vs actual a) wet biomass and b) dry biomass values.

predict wet and dry biomasses by using the selected spectral bands.

3.2. *Discussion*

The main purpose of this research was to develop the best-fit model with respect to sugarcane properties and the collected data structure to enhance the accuracy of sugarcane's wet and dry biomass estimations. Four research questions were intended to be answered in this research: 1) How can introducing proper random effects related to the sugarcane properties and the collected data structure affect the model performance? 2) Is it feasible to estimate biomass without prior knowledge of the applied nitrogen fertilizer treatments? 3) Are the extracted spectral bands efficient in estimating the sugarcane biomass; and 4) Which are the most appropriate spectral bands of multispectral images for the sugarcane biomass estimation?

The correlation between variables in the repeated measurements is high, in terms of data collection and the crop properties, which brings complexities to the development of the estimation model [29]. In comparison to other models, the mixed-effect model is a robust technique for dealing with studies on highly structured data with repeated measurements [25,60]. In this research, the linear mixed-effect model was selected to estimate wet and dry sugarcane biomasses due to its generalization capability compared to other types of mixed-effect models [61]. The model's efficiency was improved after the incorporation of the correct random effects, which depended directly on the sugarcane properties, the data structure, and objective analysis. For example, Wang et al. [21] selected the indicators for rice growth, such as cultivator types, planting patterns, and growth stages, as candidates of random effects to estimate the rice biomass. Here for addressing the first question, we examined eight separate selections of random effects along with ID, M, and T as the important indicators of the crop properties and the data structure to see what the most appropriate random effects are for correctly estimating the sugarcane growth parameters. These variables (ID, M, and T) were chosen as candidates of the random effects to cluster our data to the generally homogeneous groups. For instance, the ID variable presents the correlation for the measurements collected from the same plot. The IDM variable specifies this correlation at the row level for the samples taken from the same plot at the same growth stage. The sugarcane growth stage (M) indicates that the measurements taken from the same growth stage have a higher probability to be correlated. The results we present here indicate that for both wet and dry biomass estimations the mixed-effect models with [ID, M, IDM] and [ID, IDM] outperformed the rest of the models (Table 6). Since the models with less random effects (ID and IDM) failed to converge by the model selection strategy, we concluded the proposed models with more random effects (ID, IDM, M) are preferable in this research. Therefore, a mixed-effect model with fewer random effects is not necessarily more efficient, due to a smaller number of levels, which results in more observations per level as concluded by Wang et al. [21]. As previously mentioned, the crop properties and the data structure are critical to group observations in terms of random effects. The crop height as a

significant crop property allows researchers to investigate the development dynamics of various crop varieties under a variety of environmental conditions and agricultural management. Therefore, as a future study recommendation, Digital Surface Model (DSM)-derived crop height based on the multi-temporal UAV imagery might improve the performance of the biomass estimation model.

The importance of the second question is the nitrogen fertilizer treatment caused differences in wet and dry sugarcane biomasses and consequently in crop growth as presented in Fig. 4 and Tables 4 and 5. The biomass content at each growth stage significantly changed depending on the type of the applied nitrogen fertilizer treatment. Regarding this, introducing the T variable to the proposed model could be a necessity to estimate biomass. Generally, the T variable selection as a fixed or random effect is subject to the aim of the study. When T is treated as a fixed-effects factor, it is difficult to draw general inferences for other treatments which are not included in the model. It is only possible to compare applied T to one another and the different impacts on crop growth. On the other hand, if T is considered a random effect, the results could be generalized to all treatments. In this research, T was used in four of the random-effect selections to examine whether we could estimate sugarcane biomass efficiently using the model without utilizing the T variable. Then, we aimed at generalizing the obtained estimation models for any other type of nitrogen fertilizer treatment, in the future. Regarding the comparative analysis of the predictive capability was performed between all eight selections of random effects shown in Table 6, the proposed models excluding the T variable performed better than the models including T. Therefore, the proposed models can be considered as an efficient, sufficient, and effective way to estimate wet and dry sugarcane biomasses without prior knowledge of the applied nitrogen fertilizer. However, in the final proposed sugarcane biomass estimating models, ID and M made a significant contribution. Therefore, in order to use this model in practice, the location and growth stage of the crop must be determined first. It is simple to assign a local ID for each crop plot in a row and plot-level studies, or a local field ID for field-level estimation of crop biomass in any region/country of the world, in order to distinguish between crop locations and account for their effects in the final estimation model. The growth stage might change depending on geographical region or sugarcane variety. In such a scenario, even if the growth stage information is not accessible from ground observations records, it may be simply identified based on the time of data collection. In future studies, the model will be employed to develop estimating models at various scales, such as block or field levels, in other locations and for diverse sugarcane species.

The efficiency of the proposed models was proved by the comparative analysis in terms of R^2 and RMSE between the proposed models and three baseline models (Fig. 5 and Fig. 6). The results also addressed the third question that the UAV-based spectral bands can be successfully applied to prevent the senescence and saturation problem with the VIs and enhance the accuracies of wet and dry sugarcane biomasses. The last question of this research was addressed regarding the results of the model selection in terms of the AIC parameter. The results show the selected spectral bands can be successfully applied to wet and dry sugarcane biomasses. These selected bands not only obtained the equal accuracy compared with the mixed-effect models using all five spectral bands but also removed the redundant spectral bands. As a result, the developed models here can enhance the accuracies of the wet and dry sugarcane biomass estimations. It is likely that different results will obtain in other sugarcane growth cycles or types to those identified here. Further research with the abundance of data will tell how these models differ on other sugarcane growth cycles and types. Furthermore, the abundance of data will enable machine learning methods to be used to evaluate the performance of the suggested model here. Our intention is to pursue further research in this aspect in the future. It should be also noted that considering the nitrogen fertilizer treatments as a fixed-effects factor in the model can provide valuable information for analyzing and comparing the effects of each treatment on crop growth.

4. Conclusions

In this study, we developed the best-fit mixed-effect models in terms of fixed and random effects to improve the estimations of wet and dry sugarcane biomasses in comparison with three baseline methods (MLR, GLM, and GAM). The measurements of the plant biomass and the UAV-based multispectral imagery at different crop growth stages in the experimental sugarcane field with different nitrogen fertilizer treatments, were conducted to model the biomass estimation. The mixed-effect models outperformed the other models for both wet and dry biomass estimation, with the highest R^2 and lowest RMSE. The contributions of various random effects regarding the sugarcane properties and the collected data structure were investigated in the proposed mixed-effect models. The model selection strategy was extended to the models to analyze the most suitable spectral information as fixed effects to avoid the insufficient efficiency of the VIs in biomass estimation of a mature crop in the harvest time. The results showed that throughout the season, the Green, Blue, and NIR spectral bands correlated significantly with biomass parameters (wet and dry), while three selected VIs (NDVI, NDRE, and GNDVI) have been weakly associated with biomass parameters during the peak growth/harvest stage. In summary, the study indicated that the wet and dry sugarcane biomasses can be effectively estimated using the proposed mixed-effect models here without prior knowledge of the applied nitrogen fertilizer treatment on the crop. These findings can further serve as a useful reference for the prediction of sugarcane crop yield in various nitrogen applications.

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.

Acknowledgement

Remote sensing data were obtained from the sugarcane field trials at Bundaberg Research Facility of Queensland Department of Agriculture and Fisheries. The field trial was part of the "Smart blending of enhanced efficiency fertilizers to maximize sugarcane profitability" project under the More Profit from Nitrogen Program, supported by funding from the Australian Government Department of Agriculture, Water and the Environment as part of its Rural Research & Development for Profit program, Sugar Research Australia, and Queensland Government, with contribution from Incitec Pivot Ltd and ICL Specialty Fertilizers.

Appendix A

See Tables A1 and A2.

Table A1 – The model selection results for wet biomass.

Model Number	Intercept	Red Band	Green Band	Blue Band	NIR Band	RE Band	AIC
1	−45.13	–	−1 301.00	4 322.50	161.79	–	930.80
2	−59.42	–	−2 778.61	4 295.75	161.34	865.05	930.97
3	−46.11	274.88	−1 267.26	3 708.77	177.45	–	932.61
4	−59.47	−22.27	−2 794.93	4 345.18	160.06	873.04	932.97
5	−46.73	–	–	3 630.40	147.52	−479.50	934.26
6	32.60	–	−9 30.83	3 367.94	–	–	935.05
7	−26.68	1 340.48	−6 92.73	–	211.51	–	935.12
8	18.42	–	−2 387.85	3 353.66	–	849.65	935.25
9	−68.31	–	–	2 400.71	118.23	–	935.35
10	−46.78	513.58	–	2 575.34	178.04	−480.73	935.59
11	−51.66	991.53	–	–	170.03	–	935.61
12	−3.08	−736.15	−3 041.98	5 240.89	–	1 118.88	935.71
13	−30.06	1 285.20	–	–	204.14	−306.66	936.09
14	22.80	−4 26.93	−1 043.83	4 467.02	–	–	936.49
15	−1.55	–	–	2 077.29	–	–	936.65
16	−31.51	1 286.43	−1 370.48	–	206.45	413.34	936.73
17	−67.90	4 89.26	–	1 386.27	146.65	–	936.74
18	23.21	–	–	2 825.56	–	−311.77	937.24
19	−5.15	−108.03	–	2 317.61	–		938.62
20	17.87	−191.85	–	3 280.92	–	−324.32	939.12
21	43.57	662.42	–	–	–	–	939.29
22	61.75	764.28	−261.61	–	–	–	940.89
23	51.99	709.62		–	–	−63.52	941.22
24	51.55	707.48	−1 263.06	–	–	601.56	942.07
25	77.69	–	–	–	–	–	942.56
26	41.63	–	–	–	78.33	–	943.24
27	44.03	–	–	–	–	196.89	943.67
28	58.41	–	215.22	–	–		944.24
29	12.35	–	–	–	73.50	184.30	944.48
30	43.56	–	−1 317.33	–	–	889.95	944.51
31	26.43	–	1 85.00	–	75.34	–	945.01
32	10.61	–	−1 388.52	–	76.46	914.10	945.23

Table A2 – The model selection results for dry biomass.

Model Number	Intercept	Red Band	Green Band	Blue Band	NIR Band	RE Band	AIC
1	−10.52	–	−298.21	901.85	53.16	–	706.06
2	−16.13	217.62	–	–	57.71	–	706.17
3	−9.01	330.38	−218.74	–	72.19	–	706.30
4	−8.33	333.02	–	–	71.92	−118.65	706.52
5	−15.95	–	–	466.02	43.07		706.88
6	−9.74	–	–	796.84	50.68	−130.67	707.03
7	−11.48	177.31	−291.20	524.02	65.35	–	707.49
8	5.28	–	–	–	35.56	–	707.92
9	−11.55	–	−388.98	896.67	53.20	54.65	707.97
10	−10.28	225.38	–	352.17	66.77	−139.53	708.10
11	−16.75	194.67	–	63.12	56.39		708.15
12	−8.52	335.53	−164.67	–	72.66	−34.03	708.29
13	−11.56	175.74	−299.41	526.90	65.25	4.90	709.49
14	2.26	–	–	–	35.10	18.94	709.87
15	3.91	–	16.70	–	35.29	–	709.92
16	21.66	–	–	–	–	–	710.44
17	9.55	–	–	317.33	–	–	710.95
18	2.18	–	−124.43	–	35.54	83.43	711.73
19	4.12	−158.25	–	673.41	–	–	712.15
20	19.06	50.48	–	–	–	–	712.16
21	15.24	–	–	–	–	37.56	712.17
22	16.61	–	56.34	–	–	–	712.26
23	13.18	–	−97.19	450.36	–	–	712.64
24	12.42	–		401.58	–	−35.56	712.81
25	8.63	−210.09	−164.75	1012.12	–	–	713.32
26	7.83	−175.60	–	835.09	–	−52.52	713.84
27	15.57	34.59	–	–	–	25.17	714.07
28	16.69	39.91	32.46	–	–	–	714.11
29	15.27	–	−60.71	–	–	69.20	714.13
30	12.18	–	−187.89	445.78	–	54.44	714.55
31	5.80	−235.37	−373.81	1069.83	–	120.83	714.97
32	15.60	34.38	−59.24	–	–	56.10	716.03

REFERENCES

[1] Tian J, Wang L, Li X, Gong H, Shi C, Zhong R, et al. Comparison of UAV and WorldView-2 imagery for mapping leaf area index of mangrove forest. Int J Appl Earth Obs Geoinf 2017;61:22–31.

[2] Sofonia J, Shendryk Y, Phinn S, Roelfsema C, Kendoul F, Skocaj D. Monitoring sugarcane growth response to varying nitrogen application rates: A comparison of UAV SLAM LiDAR and photogrammetry. Int J Appl Earth Obser Geoinformat 2019;82:101878. https://doi.org/10.1016/j.jag.2019.05.011.

[3] Wang J, Badenhorst P, Phelan A, Pembleton L, Shi F, Cogan N, et al. Using sensors and unmanned aircraft systems for high-throughput phenotyping of biomass in perennial ryegrass breeding trials. Front Plant Sci 2019;10. https://doi.org/10.3389/fpls.2019.01381.

[4] Brovkina O, Zemek F, Fabiánek T. Aboveground biomass estimation with airborne hyperspectral and LiDAR data in Tesinske Beskydy Mountains. Beskydy 2015;8(1):35–46.

[5] Lumbierres M, Méndez P, Bustamante J, Soriguer R, Santamaría L. Modeling biomass production in seasonal wetlands using MODIS NDVI land surface phenology. Remote Sensing 2017;9(4):392.

[6] Morel J, Todoroff P, Bégué A, Bury A, Martiné J-F, Petit M. Toward a satellite-based system of sugarcane yield estimation and forecasting in smallholder farming conditions: A case study on Reunion Island. Remote Sensing 2014;6(7):6620–35.

[7] Pandit S, Tsuyuki S. Landscape-scale aboveground biomass estimation in buffer zone community forests of central Nepal: Coupling in situ measurements with Landsat 8 satellite data. Remote Sensing 2018;10(11):1848.

[8] Molijn RA, Iannini L, Vieira RJ, Hanssen RF. Sugarcane productivity mapping through C-band and L-band SAR and optical satellite imagery. Remote Sensing 2019;11(9):1109. https://doi.org/10.3390/rs11091109.

[9] Laneve G, Marzialetti P, Luciani R, Fusilli L, Mulianga B. Sugarcane biomass estimate based on SAR imagery: A radar systems comparison, In: 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). Fort Worth, TX, USA; 2017. p. 5834–5837.

[10] Chang A, Jung J, Maeda MM, Landivar J. Crop height monitoring with digital imagery from Unmanned Aerial System (UAS). Comput Electron Agric 2017;141:232–7.

[11] Shi L, Hu S, Zha Y. Estimation of sugarcane yield by assimilating UAV and ground measurements via ensemble Kalman filter. In: IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium. Valencia, Spain; 2018. p. 8816–8819.

[12] Som-ard J, Hossain MD, Ninsawat S, Veerachitt V. Ninsawat Sarawut, Veerachitt Vorraveerukorn. Pre-harvest sugarcane yield estimation using UAV-based RGB images and ground observation. Sugar Tech 2018;20(6):645–57.

[13] Imran AB, Khan K, Ali N, Ahmad N, Ali A, Shah K. Narrow band based and broadband derived vegetation indices using

Sentinel-2 Imagery to estimate vegetation biomass. Global J Environ Sci Manage 2020;6(1):97–108.
[14] Ampatzidis Y, Partel V. UAV-based high throughput phenotyping in citrus utilizing multispectral imaging and artificial intelligence. Remote Sensing 2019;11(4):410.
[15] Li Y, Li C, Li M, Liu Z. Influence of variable selection and forest type on forest aboveground biomass estimation using machine learning algorithms. Forests 2019;10(12):1073.
[16] Lu D. The potential and challenge of remote sensing-based biomass estimation. Int J Remote Sens 2006;27(7):1297–328.
[17] Yang S, Feng Q, Liang T, Liu B, Zhang W, Xie H. Modeling grassland above-ground biomass based on artificial neural network and remote sensing in the Three-River Headwaters Region. Remote Sens Environ 2018;204:448–55.
[18] Zhang C, Denka S, Cooper H, Mishra DR. Quantification of sawgrass marsh aboveground biomass in the coastal Everglades using object-based ensemble analysis and Landsat data. Remote Sensing of Environmen 2018;204:366–79.
[19] Han L, Yang G, Dai H, Xu B, Yang H, Feng H, et al. Modeling maize above-ground biomass based on machine learning approaches using UAV remote-sensing data. Plant methods 2019;15(1):1–19.
[20] Viljanen N, Honkavaara E, Näsi R, Hakala T, Niemeläinen O, Kaivosoja J. A novel machine learning method for estimating biomass of grass swards using a photogrammetric canopy height model, images and vegetation indices captured by a drone. Agriculture 2018;8(5):70.
[21] Wang Y, Zhang K, Tang C, Cao Q, Tian Y, Zhu Y, et al. Estimation of rice growth parameters based on linear mixed-effect model using multispectral images from fixed-wing unmanned aerial vehicles. Remote Sensing 2019;11(11):1371. https://doi.org/10.3390/rs11111371.
[22] Corti M, Cavalli D, Cabassi G, Marino Gallina P, Bechini L. Does remote and proximal optical sensing successfully estimate maize variables? A review. Eur J Agron 2018;99:37–50.
[23] Kang Y, Özdoğan M, Zipper SC, Román MO, Walker J, Hong SY, et al. How universal is the relationship between remotely sensed vegetation indices and crop leaf area index? A global assessment. Remote Sensing 2016;8(7):597.
[24] Shaver TM, Khosla R, Westfall DG. Evaluation of two crop canopy sensors for nitrogen variability determination in irrigated maize. Precis Agric 2011;12(6):892–904.
[25] Poudel K, Flewelling JW, Temesgen H. Predicting volume and biomass change from multi-temporal LiDAR sampling and remeasured field inventory data in Panther Creek Watershed, Oregon, USA. Forests 2018;9(1):28.
[26] Ou G, Wang J, Xu H, Chen K, Zheng H, Zhang B, et al. Incorporating topographic factors in nonlinear mixed-effects models for aboveground biomass of natural Simao pine in Yunnan. China J Forestry Res 2016;27(1):119–31.
[27] Bronisz K, Mehtätalo L. Seemingly unrelated mixed-effects biomass models for young silver birch stands on post-agricultural lands. Forests 2020;11(4):381. https://doi.org/10.3390/f11040381.
[28] Breslow Norman E, Clayton DG. Approximate inference in generalized linear mixed models. J Am Stat Assoc 1993;88 (421):9–25.
[29] Harrison XA, Donaldson L, Correa-Cano ME, Evans J, Fisher D, Goodwin Cecily ED, et al. A brief introduction to mixed effects modelling and multi-model inference in ecology. PeerJ 2018;6.
[30] Schober P, Vetter T. Repeated measures designs and analysis of longitudinal data: If at first you do not succeed—try, try again. Anesth Analg 2018;127(2):569–75.
[31] Costa L, Nunes L, Ampatzidis Y. A new visible band index (vNDVI) for estimating NDVI values on RGB images utilizing genetic algorithms. Comput Electron Agric 2020;172:105334. https://doi.org/10.1016/j.compag.2020.105334.
[32] Chen L, Wang Y, Ren C, Zhang B, Wang Z. Optimal combination of predictors and algorithms for forest above-ground biomass mapping from Sentinel and SRTM data. Remote Sensing 2019;11(4):414.
[33] Ballesteros R, Ortega JF, Hernández D, Moreno MÁ. Characterization of Vitis vinifera L. Canopy Using Unmanned Aerial Vehicle-Based Remote Sensing and Photogrammetry Techniques. Am J Enol Vitic 2015;66(2):120–9.
[34] Ota T, Ahmed OS, Minn ST, Khai TC, Mizoue N, Yoshida S, et al. Estimating selective logging impacts on aboveground biomass in tropical forests using digital aerial photography obtained before and after a logging event from an unmanned aerial vehicle. Forest Ecol Manage 2019;433:162–9.
[35] Sankaran S, Zhou J, Khot LR, Trapp JJ, Mndolwa E, Miklas PN. High-throughput field phenotyping in dry bean using small unmanned aerial vehicle based multispectral imagery. Comput Electron Agric 2018;151:84–92.
[36] Rahman MM, J. Robson A. A novel approach for sugarcane yield prediction using landsat time series imagery: A case study on Bundaberg region. Adv Remote Sens 2016;05 (02):93–102.
[37] Hunt ER, Hively WD, Fujikawa S, Linden D, Daughtry CS, McCarty G. Acquisition of NIR-green-blue digital photographs from unmanned aircraft for crop monitoring. Remote Sensing 2010;2(1):290–305.
[38] Amaral LR, Molin JP, Portz G, Finazzi FB, Cortinove L. Comparison of crop canopy reflectance sensors used to identify sugarcane biomass and nitrogen status. Precis Agric 2015;16(1):15–28.
[39] Abdel-Rahman EM, Ahmed FB. The application of remote sensing techniques to sugarcane (Saccharum spp. Hybrid) production: A review of the literature. Int J Remote Sens 2008;29(13):3753–67.
[40] Ryu J-H, Jeong H, Cho J. Performances of vegetation indices on paddy rice at elevated air temperature, heat stress, and herbicide damage. Remote Sensing 2020;12(16):2654.
[41] Lindsey AJ, Craft JC, Barker DJ. Modeling canopy senescence to calculate soybean maturity date using NDVI. Crop Sci 2020;60(1):172–80.
[42] Yu K, Wu X, Ding W, Pei J. Scalable and accurate online feature selection for big data. ACM Trans Knowl Discovery Data 2016;11(2):1–39.
[43] Adame-Campos RL, Ghilardi A, Gao Y, Paneque-Gálvez J, Mas J-F. Variables selection for aboveground biomass estimations using satellite data: A comparison between relative importance approach and stepwise Aakaike's information criterion. ISPRS Int J Geo-Informat 2019;8(6):245.
[44] Ratner B. Variable selection methods in regression: Ignorable problem, outing notable solution. J Targeting, Measur, Anal Market 2010;18(1):65–75.
[45] Yamashita T, Yamashita K, Kamimura R. A stepwise AIC method for variable selection in linear regression. Commun Stat-Theory Methods 2007;36(13):2395–403.
[46] Honkanen H, Kataja J. Technological aspects of nonfood agricultural lignocellulose transformations. In: Dalena F, Basile A, Rossi C, editors. Bioenergy Systems for the Future. Woodhead Publishing; 2017. p. 43–59.
[47] Shendryk Y, Davy R, Thorburn P. Integrating satellite imagery and environmental data to predict field-level cane and sugar yields in Australia using machine learning. Field Crops Research 2021;260:107984. https://doi.org/10.1016/j.fcr.2020.107984.
[48] Wang W, Reeves S. Smart blending of enhanced efficiency fertilisers to maximise sugarcane profitability. Link: 2018-MPFN_A4_QDESSugar.pdf (sugarresearch.com.au).

[49] Team RStudio. Rstudio: Integrated development environment for R, 2020.
[50] Alain Z, Ieno Elena N, Neil W, Saveliev Anatoly A, Smith GM, editors. Mixed effects models and extensions in ecology with R. Springer Science & Business Media; 2009.
[51] Avtar R, Suzuki R, Sawada H, Jose S. Natural forest biomass estimation based on plantation information using PALSAR data. PLoS ONE 2014;9(1):e86121.
[52] García-Gutiérrez J, González-Ferreiro E, Mateos-García D, Riquelme-Santos José C. A preliminary study of the suitability of deep learning to improve LiDAR-derived biomass estimation. In: International Conference on Hybrid Artificial Intelligence Systems. Switzerland; 2016. p. 588–596.
[53] Gnyp ML, Yu K, Aasen H, Yao Y, Huang S, Miao Y, et al. Analysis of Crop Reflectance for Estimating Biomass in Rice Canopies at Different Phenological StagesAnalysis of Crop Reflectance for Estimating Biomass in Rice Canopies at Different Phenological Stages. pfg 2013;2013(4):351–65.
[54] Xie Y, Sha Z, Yu M, Bai Y, Zhang L. A comparison of two models with Landsat data for estimating above ground grassland biomass in Inner Mongolia. China Ecol Modell 2009;220(15):1810–8.
[55] Xu J-X, Ma J, Tang Y-N, Wu W-X, Shao J-H, Wu W-B, et al. Estimation of sugarcane yield using a machine learning approach based on UAV-LiDAR data. Remote Sensing 2020;12 (17):2823. https://doi.org/10.3390/rs12172823.
[56] Chambers J, Hastie T, Pregibon D. Statistical models in S. In: Compstat. 1990. p. 317–321.
[57] Murase H, Nagashima H, Yonezaki S, Matsukura R, Kitakado T. Application of a Generalized Additive Model (GAM) to reveal relationships between environmental factors and distributions of pelagic fish and krill: A case study in Sendai Bay. Japan ICES J Marine Science 2009;66(6):1417–24.
[58] Wood SN. Generalized additive models: An introduction with R. CRC Press; 2017.
[59] Kamil B. MuMIn: Multi-Model Inference. Link: https://cran.r-project.org/web/packages/MuMIn/MuMIn.pdf. 2020.
[60] Chen D, Huang X, Zhang S, Sun X. Biomass modeling of larch (larix spp.) plantations in China based on the mixed model, dummy variable model, and bayesian hierarchical model. Forests 2017;8(8):268.
[61] Goldstein H. Multilevel statistical models, Vol. 922. John Wiley & Sons; 2011.