



Comparing methods to estimate perennial ryegrass biomass: canopy height and spectral vegetation indices

Gustavo Togeiro de Alckmin^{1,2} · Lammert Kooistra³ · Richard Rawnsley² · Arko Lucieer¹

Published online: 9 July 2020
© The Author(s) 2020

Abstract

Pasture management is highly dependent on accurate biomass estimation. Usually, such activity is neglected as current methods are time-consuming and frequently perceived as inaccurate. Conversely, spectral data is a promising technique to automate and improve the accuracy and precision of estimates. Historically, spectral vegetation indices have been widely adopted and large numbers have been proposed. The selection of the optimal index or satisfactory subset of indices to accurately estimate biomass is not trivial and can influence the design of new sensors. This study aimed to compare a canopy-based technique (rising plate meter) with spectral vegetation indices. It examined 97 vegetation indices and 11,026 combinations of normalized ratio indices paired with different regression techniques on 900 pasture biomass data points of perennial ryegrass (*Lolium perenne*) collected throughout a 1-year period. The analyses demonstrated that the canopy-based technique is superior to the standard normalized difference vegetation index (Δ , 115.1 kg DM ha⁻¹ RMSE), equivalent to the best performing normalized ratio index and less accurate than four selected vegetation indices deployed with different regression techniques (maximum Δ , 231.1 kg DM ha⁻¹). When employing the four selected vegetation indices, random forests was the best performing regression technique, followed by support vector machines, multivariate adaptive regression splines and linear regression. Estimate precision was improved through model stacking. In summary, this study demonstrated a series of achievable improvements in both accuracy and precision of pasture biomass estimation, while comparing different numbers of inputs and regression techniques and providing a benchmark against standard techniques of precision agriculture and pasture management.

Keywords Biomass · Perennial ryegrass · Vegetation index · Canopy height · Rising plate meter · Machine learning

✉ Gustavo Togeiro de Alckmin
gustavo.alckmin@utas.edu.au; gustavo.alckmin@wur.nl

¹ Discipline of Geography and Spatial Sciences, School of Technology, Environments and Design, University of Tasmania, Hobart, TAS 7005, Australia

² Centre for Dairy, Grains and Grazing, Tasmanian Institute of Agriculture, 16–20 Mooreville Rd, Burnie, TAS 7320, Australia

³ Laboratory of Geo-Information Science and Remote Sensing, Wageningen University, Droevendaalsesteeg 3, 6708 PB Wageningen, The Netherlands

Introduction

Efficient pasture production and utilisation is often the most critical component in a pasture-based dairy operation (García et al. 2014). Despite the extensive literature available about the constituent factors of pasture production and management, most pasture-based grazing systems are not optimally managed due to the costly and time-consuming nature of standard methods for pasture measurement and monitoring. In practice, coordinating pasture growth rates and grazing events is crucial to optimise yield (Chapman et al. 2012) while avoiding nutritional value losses (Turner et al. 2006).

In a farm scenario, such coordination can be achieved through weekly measurements of pasture biomass (García et al. 2014). Currently, common methods, such as the rising plate meter (RPM), rely on linear relationships between canopy height (Allen et al. 2011) and biomass. However, these relationships are limited in their accuracy and biased due to plant-development stages, canopy architecture (erectophile or plagiophile) or canopy density (Nakagami 2016). Finally, such a method requires a trained observer constantly monitoring and sampling paddocks, which can lead to inconsistencies due to observer-bias (Thomson et al. 1997) and requires extensive time investment (Hall et al. 2019).

Alternatively, *remote sensing* (RS) techniques can be employed to provide semi- or fully-automated monitoring, provided that the ideal temporal-spatial scales are observed. This constraint has, historically, prevented widespread adoption, as satellite revisit intervals and pixel size are not adequate for on-farm management. Furthermore, persistent cloud cover is a pervasive constraint (Kawamura et al. 2008) during seasons of rapid growth rates, hence, decreasing the utility of satellite optical imagery (Ali et al. 2016).

Unmanned aerial vehicles (UAV) and *multispectral* (MS) cameras (those defined as multi-camera 2D imagers in Aasen et al. (2018)), provide a flexible and economical system for spectral observation at a time and spatial-scale which are ideal for agricultural practices. As an advantage, such multispectral cameras are readily available, producing radiometric calibrated imagery which can be easily integrated into agricultural management tools. A drawback, however, is that multi-camera 2D imagers are mostly restricted to a small number of bands, usually five as per Aasen et al. (2018), within the *visible to near-infrared* range (VIS–NIR) of the spectrum due to their silicon-based sensors. As a consequence, end users have focused on *vegetation indices* (VI) as enhanced predictors for biomass estimation.

Vegetation indices have been, to an extent, successfully employed in *precision agriculture* (PA) and RS to estimate different biophysical and biochemical attributes of vegetation. Accordingly, a vast number of VIs have been proposed (Xue and Su 2017), surpassing the possibility of optimally or correctly utilizing each one of them without extensive domain knowledge. This issue is found in UAV remote sensing, where most commercial sensors rely on a broadband VI such as the *normalized difference vegetation index* (NDVI) to estimate a range of different, non-correlated attributes. Notwithstanding its versatility, estimations from a multi-purpose VI (e.g. NDVI) are, logically, not optimal for attributes different from which the index was originally designed. Issues such as NDVI saturation, occurring after a certain biomass threshold, are well-described in the literature (Thenkabail et al. 2002).

Despite the saturation drawback with the traditional NDVI, Mutanga and Skidmore (2004) have shown that, through systematic generation of *normalized ratio indices* (NRI) for all band pairs of a hyperspectral dataset (acquired at handheld level), certain narrow-band NRIs can overcome biomass saturation. In parallel, Burkart et al. (2014) demonstrated

the equivalence between spectral observations at two different data acquisition scales: handheld and low-level flight. In a more recent study, Wang et al. (2019) presented a brief review on the difference between reflectance measurements from both acquisition-scales, indicating that such differences are negligible. Such findings fulfil a methodological gap for data collection, analysis and performance validation for UAV sensors (i.e. low-level flight) extrapolated from handheld spectral information.

Presently, these approaches (i.e. filtering and selection of optimal NRIs and known VIs) have not been tested and translated to custom-made commercial UAV sensors. Hence, canopy based approaches, such as the RPM, are still favoured in pasture-based systems rather than spectroscopy methods despite the unique potential, at a low-cost, offered by the adoption of UAVs, multispectral sensors and a satisfactory small set of VIs.

Given this research-gap, current commercial multispectral cameras and data analysis (Michez et al. 2019) have reported poor performances for pasture biomass estimation, both when employing spectral-based ($R^2=0.35$) or canopy based (photogrammetric estimated height) techniques ($R^2=0.23$). This is in spite of the potential reported in previous research (Mutanga and Skidmore 2004) and absence of scaling-up issues (Burkart et al. 2014; Wang et al. 2019).

The research objective of this study was to evaluate how many and which VIs (i.e. *features*) should be employed for perennial ryegrass (*Lolium perenne*) biomass estimation while comparing with two traditional techniques: the rising plate meter and broadband NDVI. Consequently, accuracy improvements can be quantified based on the nature of measurements (canopy height or spectra) and number of indices. Furthermore, different regression techniques (parametric and non-parametric algorithms) were analysed, so that increments in accuracy and precision of feature(s)-algorithm pairs can be evaluated.

Methods

Experimental design

The trial was undertaken at the Tasmanian Dairy Research Facility in Elliot (TAS, Australia— $41^{\circ} 04' 57.3''$ S, $145^{\circ} 46' 21.8''$ E). The experimental layout was an array of 30 rainfed perennial ryegrass plots (dimensions of 2.0×7.5 m, with 0.35 m border at each side of the plot's longitudinal axis), arranged as two rows by 15 columns (Fig. 1).

Plots were grouped in three main blocks (10 plots per block). Each block was split in two different growth intervals: long and short or approximately 30 and 15 days, respectively. Each plot on the split-block was randomly allocated a different nitrogen (N) fertilising regime (0, 25, 50, 75 or 100 kg N ha^{-1}). The fertiliser was manually applied (i.e. top-dressing) on each plot at the start of each regrowth cycle, having urea as N source. Prior to spring (end of August) and prior to installing the experiment, phosphorus (P), potassium (K) and sulphur (S) were broadcast throughout the trial area according to soil analysis to ensure that the lack of macronutrients would not impede pasture growth.

Data collection campaigns consisted of three subsequent stages: (1) spectral measurement, (2) canopy height measurement and (3) biomass determination. In each of these stages, attention was given to minimize confounding factors and ensure independence amongst measures, given that measurements of stages (1) and (2) overlapped spatially (Fig. 2). In total, five campaigns were carried out from December 2016 to November 2017 (as per the dates of spectral measurements).

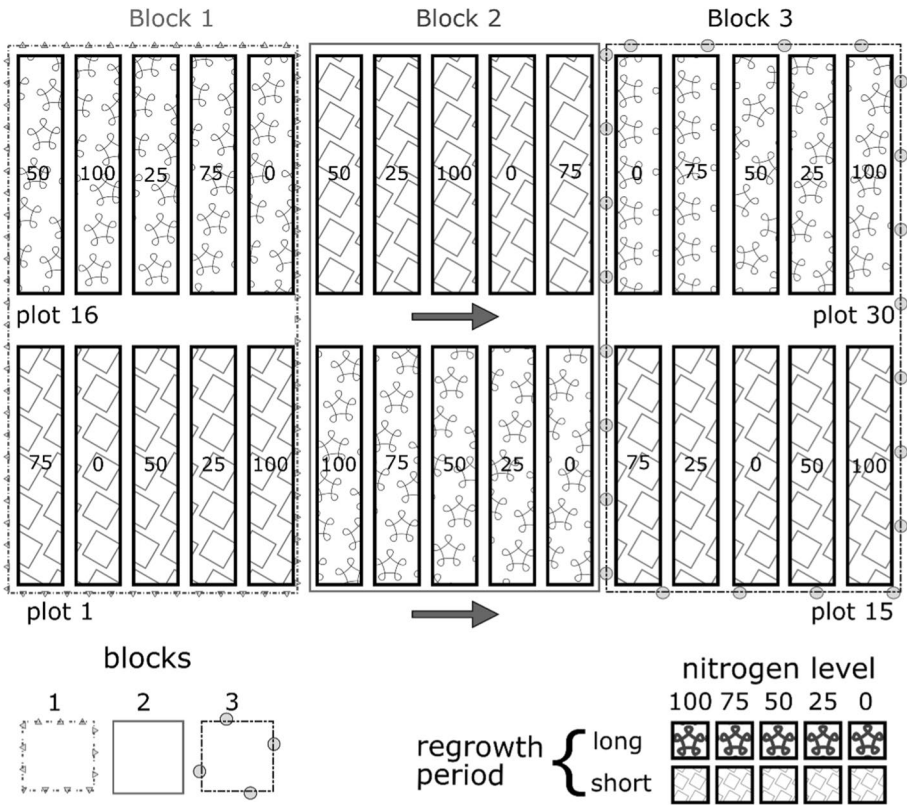


Fig. 1 Experimental layout employed in each of the five data collection campaigns. Background textures refer to regrowth periods. Nitrogen levels are indicated within the plot. Each block (encased by a dashed rectangle) comprises all combinations of regrowth periods and nitrogen fertilizing regimes. Plots are numbered from left-to-right and bottom to top. Within, each plot, six different sample-sites were measured both by the spectroradiometer and the rising plate meter

Spectral measurements

Spectral data was collected by a field spectroradiometer (ASD Hand-held 2, Colorado, USA) on five dates under clear-sky conditions and around solar noon: December 18th, 2016, February 06th, April 29th, October 22nd and November 28th, 2017. This instrument acquires data from 325 to 1075 nm, with a total of 750 bands and field of view (FOV) of 25°. Total time spent to obtain all spectral measurements (180 data points) was 1.5 to 2 h per field campaign with minimum warm-up of 30 min. The instrument setup follows the manufacturer’s recommendation: 30 scans for spectrum averaging, 60 scans for dark current and white reference. The sequence of measured plots was randomized to minimise any systematic effect of solar position across the plots during data collection. In addition, after finishing measuring the samples of each plot, a spectral measurement of the white reference (Spectralon®) was recorded. The intention of this procedure was twofold: (a) to monitor the stability of the instrument and (b) detect any possible change in atmospheric conditions. The instrument was recalibrated (against the white reference) after seven minutes of continuous usage or whenever the white-reference measurement deviated from

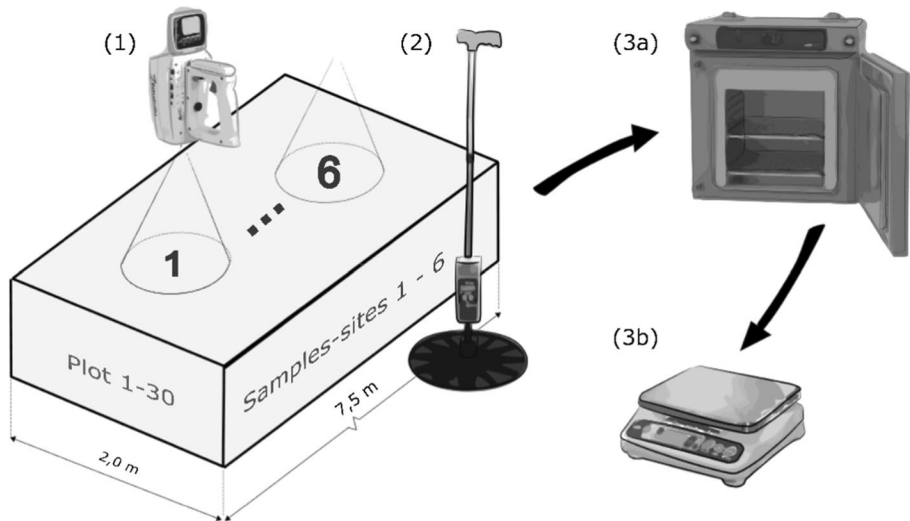


Fig. 2 Data collection workflow. Stages: (1) spectral measurements (spectroradiometer); (2) canopy height measurements (rising plate meter); 3 biomass determination, (a) drying (forced-air oven), (b) weighing (digital scale). Plots (1–30) dimensions: 2.0×7.5 m. Sample-sites (1–6) were measured both with the spectroradiometer and the rising plate meter

100% reflectance, whichever occurred first. Within each plot, six randomly allocated sample-sites were selected (Fig. 2 (1)). Spectral measurements were taken from approximately one-meter height, thus, yielding a circular footprint equal to 0.15 m² (or 0.44 m diameter). Each sample-site was measured five times. Final sample spectral value (referred to as raw spectral data) was the average value of these five measurements.

Canopy height measurements

An analogue RPM with 5 mm resolution, described in Earle and McGowan (1979), was employed to measure (compressed) canopy height, once per sample-site (Fig. 2 (2)).

Biomass determination

Pasture biomass was mechanically defoliated above a residual height of 50 mm from the 0.15 m² footprint used in stage 1 and 2 (Fig. 2 (1) and (2)). Harvested material was dried for a minimum of 48 h at 60 °C in a forced-air oven (Fig. 2 (3a)) immediately following each harvest for pasture *dry-matter* (DM) determination. Samples were weighed (Fig. 2 (3b)) using a digital scale (MassCal, 30 kg ± 0.5 g).

Data analysis

For reproducibility purposes, data analysis operations are introduced by the corresponding *package::function* format (italics typeface and accompanied by the double colon operator, i.e. the *scope resolution operator*).

Data analysis was performed in RStudio/R (versions 1.14 and 3.5.1, respectively). Necessary packages for the analysis, besides the base and dependencies packages, are *hsdar* (Lehnert et al. 2018), *caret* (Kuhn 2008) and *caretEnsemble* (Mayer and Knowles 2015) (Fig. 3).

Feature generation

The raw spectra data were smoothed using the Savitzky–Golay filter (*hsdar::smoothSpeclib*, window size = 9 nm, polynomial-degree = 2). This operation aims to decrease the influence of random instrument noise without distorting the original reflectance values. As a consequence, this operation improves the signal-to-noise ratio.

Vegetation indices

The smoothed spectra were transformed into a set of 97 VIs (*hsdar::vegindex*) available in the literature. The full list of VIs can be found in Lehnert et al. (2018) and are listed in Table 2. As a baseline approach and as a complementary step for data analysis, all VIs were fitted in a univariate (single-index) ordinary linear regression against biomass values.

Normalized ratio indices

The smoothed resampled spectra were used in an exhaustive-search process, testing all available bands combinations in a normalized difference equation, as described in Mutanga and Skidmore (2004) and shown in Eq. 1. The smoothed spectra were resampled from 750 to 149 bands given that the necessary computation time to test all band combinations would be substantial, and most likely superfluous due to high spectral band correlation. Spectra were resampled (*hsdar::spectralResampling*) applying a gaussian response function and 10 nm bandwidth, reducing computational load to 4% of all possible combinations. In total, 11,026 combinations were assessed.

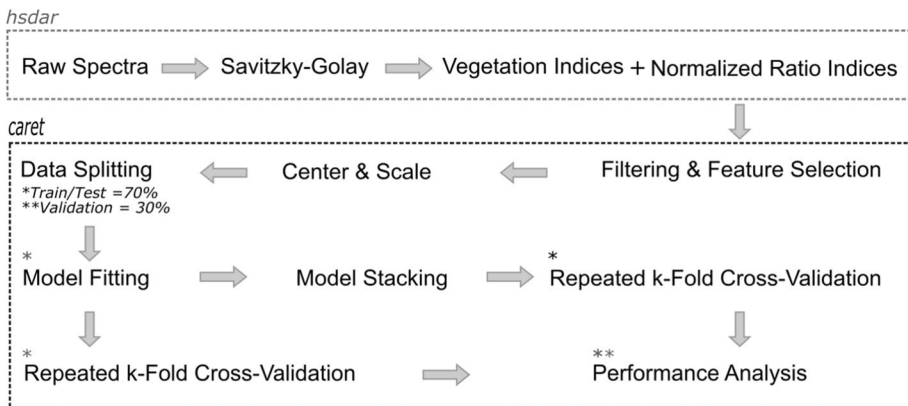


Fig. 3 Data analysis workflow. Statistical packages employed are indicated on the top-left of each dashed rectangle. Feature generation process is enclosed by the top dashed rectangle while feature filtering, model fitting and performance analysis are encased in the lower dashed rectangle. Single and double asterisks indicate where train/test and validation dataset were, respectively, employed

$$\text{NRI}(\lambda_1, \lambda_2) = \frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2} \quad (1)$$

All the NRIs were then fitted in a univariate (single-index) ordinary linear regression (*stats::lm*) against biomass values. From these, the best fitting linear model (*hdsar::nri best performance*, highest R^2) was identified and the best performing NRI (λ_1, λ_2 —*optimised NRI*) included in the pool of filtered VIs.

Feature filtering and selection

The framework for data analysis consisted of three steps: (A) *filtering* of highly correlated and non-significant VIs (i.e. features); (B) *recursive feature elimination* and feature selection (optimal and satisfactory subsets) and (C) *model fitting, stacking and validation*. A detailed description and rationale of the feature filtering and selection workflow is discussed in Perez-Riverol et al. (2017) and a specific case-study for perennial ryegrass is available in Alckmin et al. (2019). The training/testing set ($n=630$) comprises 70% of the entire dataset and was used in a repeated k -fold cross-validation (folds = 10 and repeats = 5).

Filtering

Pearson correlation among all VIs was calculated. A maximum cut-off of |0.95|, based on a sensitivity analysis developed in Alckmin et al. (2019), was applied to identify highly correlated VIs. Such VIs were evaluated in a pair-wise fashion: the one with the largest mean correlation (i.e. correlation with all other features) was removed. A minimum Pearson correlation removal cut-off equal to |0.2| between the remaining filtered VIs and DM values was applied.

Recursive feature elimination After the filtering process, remaining VIs were centred, scaled and a recursive feature elimination process (*caret::rfe*) was performed using the training set. Two subsets were identified: optimal (minimal RMSE) and the satisfactory subset, which was the smallest group of features that presented results (in training–testing stages) which were below a 10% threshold from the minimum RMSE model (optimal subset). This workflow and guidelines were presented in Kuhn and Johnson (2013) and Perez-Riverol et al. (2017).

Ranking of VIs (variable importance, *caret::varImp*) for each different VI subset size, at each cross-validation fold, was calculated through a random forests routine using Gini Importance, introduced in Breiman et al. (2017).

Model fitting, stacking and validation After the satisfactory feature set was determined (i.e. *Selected VIs*), different regression algorithms were fitted to the data. The following models were chosen: “*Bagged MARS*” (Friedman 1991), “*Random Forests*” (Breiman 2001) and “*Support Vector Machines with Polynomial Kernels*” (Cortes and Vapnik 1995) and “*Ordinary Linear Regression*” (referred to hereafter as MARS, RF, SVM and LM, respectively). The rationale for this selection was to test models that do not share the same core technique or are simply variations of the same technique. Tuning of hyper-parameters was performed automatically at the training-test stage (*caret::train*) through an embedded grid-search algorithm. Model ensemble/stacking consisted of training (*caretEnsemble*) an algorithm to combine the predictions of other previously trained algorithms (Caruana et al.

2004). Such an approach was generated through a generalized linear model of the three non-parametric models (referred to as *STACK*).

Finally, models were validated against an unseen dataset, corresponding to 30% of total observations ($n=270$), where it is expected that model performance is similar (or superior) to the training-test stage. For benchmarking purposes, *RPM*, *NDVI*, *optimised NRI* and the satisfactory subset (i.e. *selected VIs*) were all fitted using a linear model and underwent the *Performance Analysis* protocol, allowing a comparison between features.

Performance analysis

In this study, the error of each feature(s)-algorithm pair, or the difference between predicted and observed/true values (sample weight), was assessed using three different metrics: *root-mean-square error* (RMSE), *coefficient of determination* (R^2) and *mean absolute error* (MAE). Ultimately, this analysis aimed to (i) assess the accuracy (Joint Committee for Guides in Metrology 2008) for each feature(s)-algorithm pair; (ii) to test if the accuracy differences between the feature(s)-algorithm pairs were statistically significant (p -value > 0.05); (iii) if the error-distribution (i.e. precision) of these feature(s)-algorithm pairs were statistically different (Fig. 4).

For items (i) and (ii), the algorithm (*caret::diff.resamples*) and workflow provided in Kuhn and Johnson (2013) was employed, which are descriptions of benchmark studies mostly based in one sample t-test. To check if the error-distribution was different between feature(s)-algorithm pairs, item (iii), the Kolmogorov-Smirnov test (Massey 1951) was employed. The two-sample Kolmogorov-Smirnov (*KS*) test is a non-parametric analysis that compares the cumulative distributions of two datasets. In this context, it was employed using error metrics (RMSE) derived from each feature(s)-algorithm pair. The error-metrics ($n=50$) were generated through a repeated k -fold cross-validation (5 folds, 10 repeats) on the training set. Feature(s)-algorithm pairs are presented in Fig. 4 under the “*Model Fitting*” section. A schematic diagram of this workflow is presented in Fig. 4.

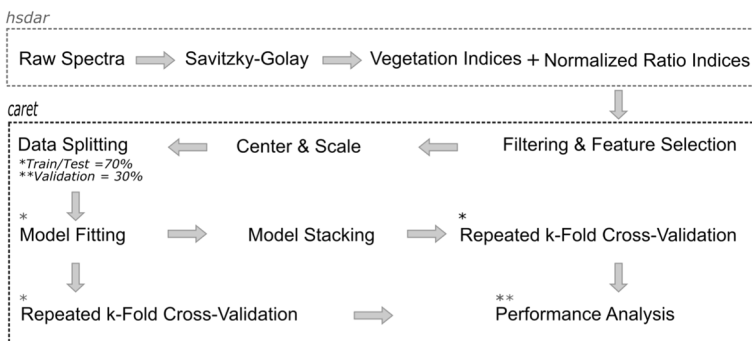


Fig. 4 Feature(s)-algorithm pair model fitting (left-side), performance and tests workflow (right-side). Selected VIs were used as features on non-parametric models. Non-parametric models were used as features on the *STACK* model

Results

Experimental setup—biomass and canopy height

Pasture biomasses ranged from 164 kg DM ha⁻¹ (minimum) to 4663 kg DM ha⁻¹ (maximum) and a mean of 1633 kg DM ha⁻¹. The descriptive statistics data of biomass and canopy height values, per campaign, can be found in Table 1.

Feature generation

Vegetation indices

The result of a baseline approach (i.e. single index linear regression), and consequent model fit metrics, can be examined in Table 2. The optimised NRI (“*Opt NRI*”), retrieved within this analysis ranks as the best performing VI (Table 2) as found by Mutanga and Skidmore (2004).

Optimised normalized ratio indices

Figure 5 presents the R² value for each NRI (λ_1, λ_2 combination) and pasture biomass values. The region with the highest correlation to pasture biomass (delimited by a dashed rectangle) occurred within the far end of the red-edge region (680–730 nm) until the NIR shoulder (up to 900 nm). The best performing NRI ($\lambda_1 = 745, \lambda_2 = 755$ nm) is then added to the feature-space in which the feature selection will be performed.

Feature filtering and selection

The filtering selection protocol reduced the number of features from 97 to 19 VIs. Within the selection workflow, the optimal (filled circle) and satisfactory (filled triangle) features subset were identified in the feature selection protocol (Fig. 6). By this protocol, the number of VIs necessary to fulfil this constraint (i.e. 10% threshold above minimum RMSE) was equal to four and are listed in Table 3 as well as indicated in boldface in Table 2. Complementary, the optimal model, or the model with lowest RMSE value, includes all 20 VIs (i.e. filtered VIs plus the optimized NRI). However, the increments in performance are marginal from six features onward (Fig. 6).

Table 1 Descriptive statistics for biophysical characteristics per campaign

Date	Canopy height (mm)				Biomass (kg DM ha ⁻¹)			
	Average	SD	Max	Min	Average	SD	Max	Min
Dec-16	124.2	31.1	225	75	1837.3	677.0	4222.2	322.3
Feb-17	65.2	20.3	115	30	1115.5	829.8	3801.3	164.4
Apr-17	56.8	19.8	115	20	1117.4	471.4	2380.7	269.6
Oct-17	129.8	27.8	205	50	2450.1	943.2	4662.8	559.0
Nov-17	99.2	37.7	210	45	1646.2	877.4	3761.8	401.2
Total	95.1	41.0	225	20	1633.3	923.0	4662.8	164.4

Table 2 Performance by vegetation index in an ordinary linear regression

Vegetation Index	R ²	RMSE	Vegetation Index	R ²	RMSE
<i>Opt NRI</i>	0.64	530.3	PARS	0.46	651.5
RDVI	0.63	538.7	SR2	0.45	654.3
SAVI	0.63	541.9	<i>Datt</i>	<i>0.45</i>	<i>654.6</i>
MSAVI	0.61	553.8	mNDVI	0.45	657.7
OSAVI	0.60	557.0	TCARI2/OSAVI2	0.45	658.7
OSAVI2	0.60	557.3	SR	0.44	660.4
<i>DD</i>	<i>0.60</i>	<i>559.5</i>	<i>Vogelmann3</i>	<i>0.44</i>	<i>661.5</i>
MCARI2/OSAVI2	0.60	562.4	PRI norm	0.43	668.5
Vogelmann2	0.59	564.9	PSRI	0.42	676.9
Vogelmann4	0.59	570.4	PRI	0.41	681.5
MCARI2	0.59	570.7	SIPI	0.40	685.8
Datt2	0.57	578.1	NPCI	0.40	688.4
MTVI	0.57	579.4	SRPI	0.39	694.7
Sum Dr2	0.57	583.2	SR7	0.38	695.1
SPVI	0.56	588.7	Datt5	0.38	698.0
TVI	0.56	588.9	GDVI 2	0.38	698.7
Sum Dr1	0.56	589.9	NDVI3	0.37	704.2
Vogelmann	0.55	591.9	SR5	0.36	707.9
Carter4	0.55	592.5	<i>Datt4</i>	<i>0.35</i>	<i>714.2</i>
Green NDVI	0.55	596.0	DWSI4	0.33	727.9
SR6	0.54	597.8	GI	0.32	729.1
NDVI2	0.54	600.5	SR4	0.32	731.4
mND705	0.54	600.6	<i>Boochs</i>	<i>0.31</i>	<i>733.7</i>
mSR705	0.54	602.9	<i>MCARI</i>	<i>0.31</i>	<i>736.6</i>
Boochs2	0.54	603.9	SR8	0.29	745.3
REP Li	0.53	604.3	Carter5	0.29	745.7
mSR2	0.53	606.8	GDVI 3	0.28	753.0
D1	0.53	607.4	TCARI2	0.28	754.1
EGFR	0.53	607.9	Datt6	0.25	766.4
MTCI	0.53	609.7	MPRI	0.24	770.3
GMI1	0.52	612.0	CI	0.24	773.5
SR3	0.52	612.0	Carter	0.23	777.9
CI2	0.52	613.2	GDVI 4	0.20	790.8
EGFN	0.51	617.7	TGI	0.17	806.6
PWI	0.51	620.4	TCARI	0.17	808.5
GMI2	0.51	621.1	CRI1	0.11	834.2
SR1	0.51	621.1	mSR	0.10	842.7
Carter3	0.50	623.5	MCARI/OSAVI	0.09	843.6
Gitelson2	0.50	625.5	Gitelson	0.09	845.6
Carter2	0.50	626.9	CIInt	0.07	854.3
DDn	0.49	629.8	CRI2	0.06	860.4
Maccioni	0.49	630.6	DPI	0.05	864.9
PSND	0.49	635.1	TCARI/OSAVI	0.04	867.6
mREIP	0.48	637.8	Carter6	0.03	874.1
Datt3	0.48	638.8	EVI	0.02	876.4
NDVI	0.48	640.3	CRI3	0.01	880.6

Table 2 (continued)

Vegetation Index	R ²	RMSE	Vegetation Index	R ²	RMSE
REP LE	0.48	641.6	PRI×CI2	0.01	880.6
PSSR	0.47	645.9	CRI4	0.01	881.0
D2	0.46	649.0	CARI	0.00	886.0

Bold italic indices are selected within the satisfactory subset. *Italic indices* are those remaining after the filtering process

Obs Values corresponding to an ordinary linear regression (*stats::lm*)

Fig. 5 Normalized Ratio Index Correlogram (λ_1 , λ_2 and biomass). Displaying the determination coefficients between biomass and narrow band NRI values calculated from all possible combinations spread across λ_1 (350–1075 nm) and λ_2 (350–1075 nm). Dashed rectangle limits λ_1 700–900 and λ_2 700–760 nm

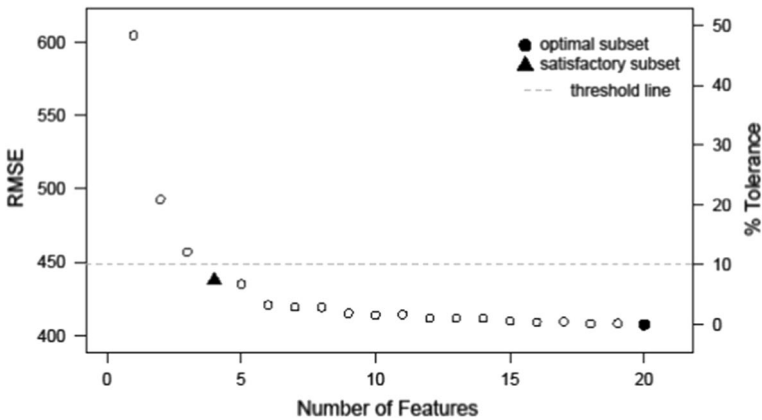
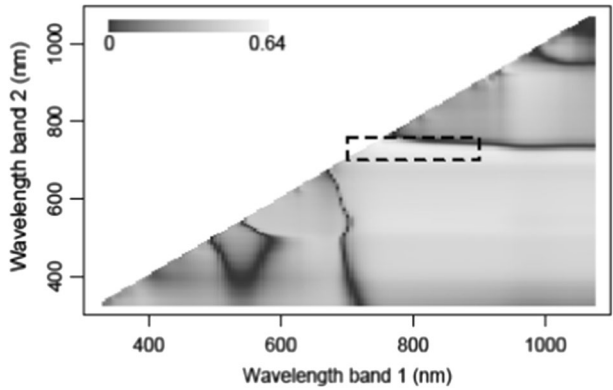


Fig. 6 Feature selection protocol. Average RMSE by number of VIs (features). Optimal (filled circle) and satisfactory (filled triangle) subsets have 20 and four VIs, respectively. Optimal subset corresponds to the lowest average RMSE found through the feature selection process. The satisfactory subset corresponds to the minimal number of features (i.e. VIs) below a ten percent tolerance threshold over the minimal RMSE

Variable importance

The four selected VIs (*satisfactory subset*) used as features for different regression techniques are: the *optimised NRI*, *Chlorophyll Index*, *Simple Ratio 7* and *Double Difference*

Table 3 Selected vegetation indices (satisfactory subset), variable importance, formulas and references

Vegetation index	Importance	Formula	Author (year)
Optimized NRI	29.7	$(\lambda_{755} - \lambda_{745}) / (\lambda_{755} + \lambda_{745})$	Mutanga and Skidmore (2004)
chlorophyll index	27.5	$(\lambda_{675} \times \lambda_{690}) / \lambda_{683}^2$	Zarco-Tejada et al. (2003)
Double difference, n	17.2	$2 \times (\lambda_{710} - \lambda_{660} - \lambda_{760})$	Lichtenthaler et al. (1996)
Simple ratio 7	14.6	$\lambda_{440} / \lambda_{690}$	Maire et al. (2008)

near the red-edge. Variable importance, extracted as per the description in Methods, respective formulas and bibliographical references are listed in Table 3.

Performance analysis

From the repeated *k*-fold cross-validation, the error metrics for each feature(s)-algorithm pair are summarised in Table 4. It presents, in descending order, the feature(s)-algorithm pair performances. Results range from (max) 770.6 to (min) 325.3 kg DM ha⁻¹ (RMSE) from NDVI and SVM, respectively. In terms of precision, the *standard-deviation* (SD) ranges from 59.5 to 15.6 kg DM ha⁻¹ (RMSE).

The performance of RF and STACK models are equivalent in terms of average accuracy. However, the STACK model is more precise with a standard-deviation of 15.6 kg DM ha⁻¹ in contrast to 41.7 kg DM ha⁻¹ of the RF model. Such difference in error-distribution is statistically significant (*p*-value < 0.01) as per the results of a two-sample KS-test results, displaying narrower error-distribution for the STACK model (Fig. 7b).

Figure 7 presents the density plot of error-distribution of each feature(s)-algorithm pair, from which a visual assessment of accuracy and precision can be extracted. This figure is divided into a density plot where only an ordinary linear regression (parametric) model was employed (Fig. 7a) and models where different non-parametric models and the selected VIs were employed (Fig. 7b).

The visual analysis provided a clear indication of the higher precision of the STACK model in comparison with other methods. In the interest of avoiding similar figures, only

Table 4 Posterior analysis results for each feature(s)-algorithm pair

Models	RMSE				R ²				MAE			
	Average	SD	Max	Min	Average	SD	Max	Min	Average	SD	Max	Min
NDVI	636.9	59.5	770.6	532.1	0.49	0.05	0.61	0.39	495.0	38.1	567.5	430.0
OptNRI	520.3	51.0	692.2	406.2	0.65	0.08	0.77	0.45	412.0	36.2	514.1	311.1
RPM	521.8	58.8	645.5	374.8	0.68	0.07	0.82	0.54	406.6	45.8	511.8	296.8
LM	475.8	46.0	571.9	354.9	0.71	0.06	0.84	0.55	371.5	36.4	442.3	295.7
MARS	453.8	44.8	528.5	338.5	0.74	0.06	0.86	0.56	358.3	36.6	419.2	279.6
SVM	425.7	45.1	513.4	325.3	0.77	0.06	0.88	0.61	332.6	34.5	401.1	259.2
RF	407.6	41.7	493.5	327.6	0.79	0.05	0.88	0.63	312.3	33.1	386.7	248.9
STACK	405.8	15.6	434.2	367.5	0.79	0.02	0.82	0.74	310.7	12.5	337.8	278.3

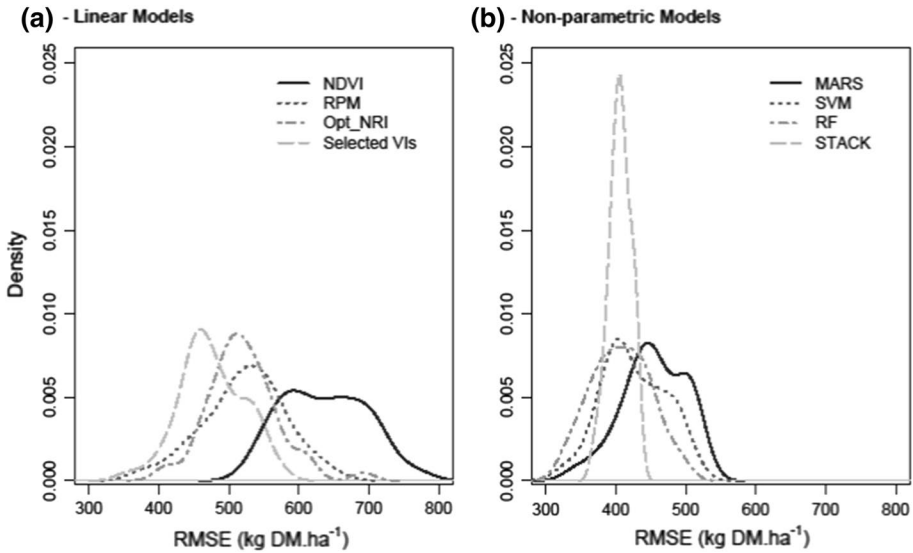


Fig. 7 Density plot of the error-distribution RMSE for each model **a** different features fitted to a linear model. **b** Selected VIs (features) fitted using non-parametric methods

the RMSE density plot will be presented as the same information for MAE and R² are also presented in Fig. 8. The ranking and performance (average error metrics and error-distribution) of each feature(s)-algorithm pair is presented in Fig. 8.

The results of the benchmark methodology developed by Kuhn and Johnson (2013) are presented in Table 5. With regard to accuracy differences, the best performing feature(s)-algorithm pair is the STACK of models with an average RMSE = 405.8 and MAE = 310.7 kg DM ha⁻¹. The lowest performing feature-algorithm pair is NDVI based on an ordinary linear model. Likewise, the saturation of NDVI becomes noticeable at

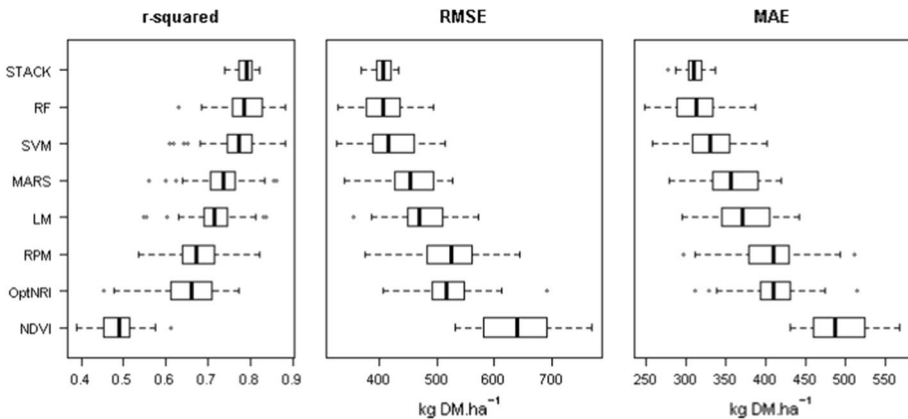


Fig. 8 Error distribution presented through boxplots. The four top boxplots refer to non-parametric models. Conversely, the four boxplots on the bottom relate to ordinary linear models. The STACK model presents a narrower error-distribution

Table 5 Posterior Analysis—RMSE, MAE and R² for feature(s)-algorithm pairs

Models	MARS	LM	SVM	RF	OptNRI	RPM	NDVI	STACK
MAE								
MARS		-25.10	11.36	31.52	-69.12	-59.11	-147.45	36.83
LM	0.00		36.46	56.61	-44.02	-34.01	-122.35	61.92
SVM	0.01	0.00		20.15	-80.48	-70.47	-158.81	25.46
RF	0.00	0.00	0.00		-100.63	-90.63	-178.96	5.31
OptN RI	0.00	0.00	0.00	0.00		10.01	-78.33	105.94
RPM	0.00	0.00	0.00	0.00	1.00		-88.34	95.94
NDVI	0.00	0.00	0.00	0.00	0.00	0.00		184.27
STACK	0.00	0.00	0.00	1.00	0.00	0.00	0.00	
RMSE								
MARS		-38.31	13.31	29.26	-78.45	-74.40	-189.51	41.60
LM	0.00		51.62	67.57	-40.15	-36.09	-151.21	79.91
SVM	0.00	0.00		15.95	-91.76	-87.71	-202.82	28.29
RF	0.00	0.00	0.00		-107.72	-103.66	-218.78	12.34
OptN RI	0.00	0.00	0.00	0.00		4.05	-111.06	120.05
RPM	0.00	0.02	0.00	0.00	1.00		-115.11	116.00
NDVI	0.00	0.00	0.00	0.00	0.00	0.00		231.11
STACK	0.00	0.00	0.00	1.00	0.00	0.00	0.00	
R ²								
MARS		0.04	-0.02	-0.03	0.10	0.06	0.25	-0.05
LM	0.00		-0.06	-0.08	0.05	0.02	0.21	-0.09
SVM	0.00	0.00		-0.02	0.11	0.08	0.27	-0.03
RF	0.00	0.00	0.01		0.13	0.09	0.28	-0.02
OptN RI	0.00	0.00	0.00	0.00		-0.04	0.15	-0.15
RPM	0.00	1.00	0.00	0.00	0.24		0.19	-0.11
NDVI	0.00	0.00	0.00	0.00	0.00	0.00		-0.30
STACK	0.00	0.00	0.01	1.00	0.00	0.00	0.00	

Upper diagonal: average difference between each feature(s)-algorithm pair (MAE, RMSE are expressed in kg DM ha⁻¹). Lower diagonal: *t*-test *p*-value for H0: difference = 0.05. *p*-value adjustment: bonferroni

around 2.500 kg DM ha⁻¹, whereas the optimised NRI does not present the same pattern (Fig. 9). As a consequence, the maximum RMSE difference between feature(s)-algorithm pairs is between the STACK models and NDVI (231.1 kg DM ha⁻¹). Additionally, the optimised NRI has, on average, a Δ (difference) in RMSE of 111.1 kg DM ha⁻¹ smaller than the NDVI (Table 5).

Concerning error-distribution (precision), most noticeably the results of the *t*-tests (lower diagonal—Table 5), indicate that the average inaccuracy (RMSE and MAE) of (i) the STACK and RF models and (ii) RPM and optimised NRI are not statistically different (Table 5). Given such, it follows to check if each of these two pairs, (i) RPM/optimised NRI and (ii) RF/STACK, share equivalent error-distributions.

The *p*-value (0.717) for the two-sample KS-test indicates that the error-distribution between the RPM and the optimised NRI are not statistically different. These results

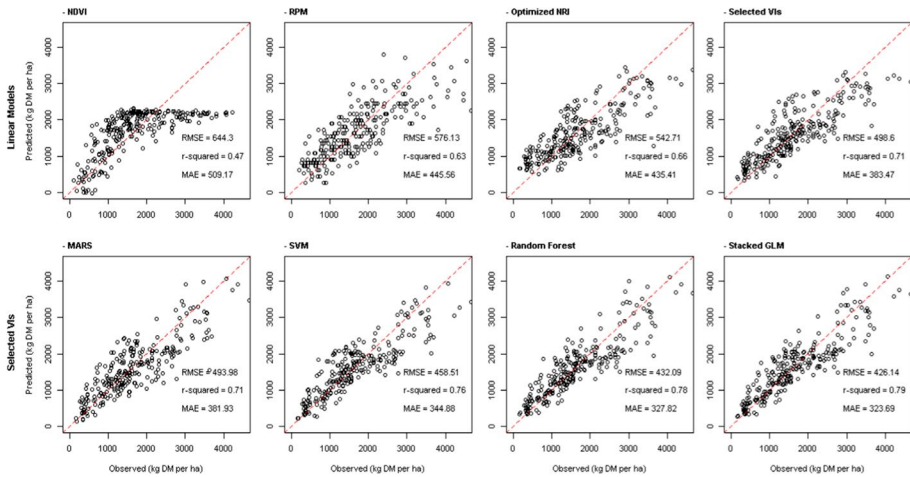


Fig. 9 Scatterplots of predicted vs observed data of all feature(s)-algorithm pairs on the validation dataset ($n=270$). The first row presents different features using an ordinary linear regression (parametric-regression). The second row uses the four selected VIs (i.e. *satisfactory subset*) in different (non-parametric) regression algorithms. Noticeably, the NDVI (top-left) scatterplot illustrates its saturation issue at around $2.500 \text{ kg DM ha}^{-1}$. Scatterplots have been arranged according to accuracy improvements and follow a left-to-right and top-to-bottom order

indicates that both accuracy (Table 5—*t*-test diagonal RMSE) and precision (KS-test) of the RPM and optimised NRI are equivalent.

In contrast, for the STACK and the RF algorithm, the KS-test *p*-value was 0.01. That indicates that both (error) distributions are unlikely ($p\text{-value} < 0.05$) to share the same distribution. In practical terms, these *post-hoc* tests indicate that the use of the STACK model provides more precise estimates, making it a more stable model, given that the standard-deviation of the error metrics (Table 4) for the STACK model is smaller than any other feature(s)-algorithm pair (Fig. 8).

Validation results, in which trained-tested models are applied to an unseen (validation) dataset, are presented in Fig. 9. The first row of plots presents predicted vs. observed results for different features in ordinary linear regressions (parametric-model). The second row presents the performance of selected VIs in different non-parametric regression algorithms.

Discussion

This study successfully selected and validated a subset of (four) VIs and regression technique (STACK) that largely outperforms the conventional method (RPM) for biomass estimation by an average ΔRMSE of $116 \text{ kg DM ha}^{-1}$ (Table 5).

This analysis has also shown that, in comparison to the most commonly adopted method on precision agriculture (i.e. standard NDVI), the best method substantially decreases the average inaccuracy and error-distribution ($\text{RMSE} \pm \text{SD}$) of biomass estimation from 637 ± 60 to $406 \pm 16 \text{ kg DM ha}^{-1}$ (Table 4), while eliminating its characteristic saturation issue (Fig. 9—NDVI). In retrospect, given the worse performance of NDVI in comparison with the RPM ($\Delta \text{RMSE} = 115 \text{ kg DM ha}^{-1}$), it is clear why end users have not adopted NDVI over RPM measurements.

The approach of an exhaustive-search of an optimised NRI identified a band combination that, both in terms of accuracy and precision, is equivalent to the RPM (Table 5). The performance of this particular VI (optimised NRI $\lambda_1=745$, $\lambda_2=755$ nm) is in-line with the results reported by Mutanga and Skidmore (2004). Equivalent performance and similar spectral band combinations were also presented in Cho et al. (2007), who reported that such combination overcomes the saturation issue presented by the standard NDVI (Fig. 9—optimised NRI).

An additional set of 97 other VIs was tested and compared based on a linear regression performance (Table 2): all were suboptimal in performance when compared to the optimised NRI. However, it is important to remark that the performances of these 97 VIs as predictors may be handicapped by a poor fit (Table 2), as the best fit function between a VI and biomass is not necessarily linear. Yet, the analysis demonstrated that the VIs extracted from the literature present a high degree of multicollinearity. The process of filtering eliminated 80.4% (78 out of 97) of features (VIs): a strong indication of redundancy and limitation of VIs. Also, the selection protocol indicated that models with more than four VIs only yield marginal accuracy improvements (Fig. 6).

This analysis also quantified the improvement in error-metrics when employing non-parametric models in comparison to a parametric model (i.e. ordinary linear regression—LM). On average, the use of ordinary linear regression, with the same selected VIs, presented an additional 38.3, 51.6, 67.6 and 79.9 kg DM ha⁻¹ error (RMSE) compared to MARS, SVM, RF and STACK models, respectively. It is important to highlight that such improvement in performance is due exclusively to model selection and tuning, improving results without requiring additional features.

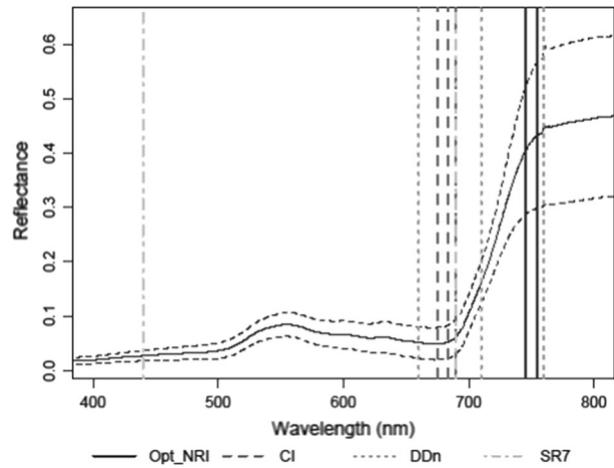
While there are accuracy improvements based solely on the choice of employing either parametric or non-parametric models (e.g. STACK—LM= Δ 79.9 kg DM ha⁻¹), such improvements are not so substantial within the group of non-parametric models. The largest difference between non-parametric models (RF—MARS) was equal to 29.3 kg DM ha⁻¹. Overall, the largest improvements in performance are due to a better selection of features (e.g. LM—NDVI= Δ -151.2) rather than a better performing regression algorithm. This highlights the importance of feature selection and feature engineering (e.g. optimised NRI process).

It was not surprising that both the RF and STACK techniques share the same average accuracy. Random forests employs the technique of bootstrapping the dataset and aggregating results (known as bagging). The same core concept is used in stacked models. However, STACK models present a better precision level, given that their output is an ensemble of other models (not features).

In essence, this analysis has shown that the use of a small feature set of VIs coupled with non-parametric methods enhances the accuracy and precision of pasture biomass estimation. Such VIs are located in the region of the red-edge and NIR-shoulder (Fig. 10), regions which have been identified as important for chlorophyll and/or leaf area index (LAI) assessment (Darvishzadeh et al. 2008; Delegido et al. 2011) as well as for pasture biomass estimation (Clevers et al. 2007).

It is reasonable to hypothesize that the selection of VIs around the red-edge could be linked with an underlying physiological process as canopies with higher levels of biomass will present higher LAI and chlorophyll mass (drivers of shifts on the red-edge). Previous studies have reported similar findings: Horler et al. (1983) credits the red-edge shift, centred at around 740 nm, to leaf stacking, thus, showing a causal link with biomass. Tucker (1977) and Collins (1978), indicated the same spectral region (740 nm) and behaviour when estimating attributes linked to biomass. Guyot and Baret (1988) examined the effects of higher chlorophyll content

Fig. 10 Selected VIs through the feature selection workflow. In the background, the mean (solid line), upper and lower boundaries (\pm SD, dashed line) of all spectral observations. Different line styles identify each selected VI. All VIs have at least one band within the regions of chlorophyll absorption, the red-edge and the NIR-shoulder



and LAI on the spectral behaviour around the red-edge. Finally, Mokvist et al. (2014) discussed a wider range of the limits of the far-red spectral absorption, demonstrating a plasticity for light absorption at higher biomass levels, indicating a characteristic which might overcome saturation.

Given the similar results and the mechanistic link explored by these previous studies, the approach and findings become more robust and, most likely, generalisable to other locations or contexts.

The plateau of 430 kg DM ha⁻¹ (RMSE) is equivalent to approximately 6.5 g per sample area (0.15 m²). Thus, a fraction of this error may well be due to the sampling/harvesting technique or even a drying process that, despite best efforts, has an inherent random noise. It is reasonable, therefore, to assume that this methodology was able to explain most of the variance which could be mapped by VIs. Further performance improvements in error metrics are mostly likely due to overfit or noise-fitting. As illustrated in Fig. 6, the addition of extra VIs will only yield marginal improvements. Under normal field-conditions, it seems unlikely that measurement error could be decreased further.

The ranges presented in Table 1 are in-line with growth rates, canopy architecture and height displayed in different seasons as per the intrinsic ecophysiological characteristics of cool-season grasses (Christie et al. 2018). Given the significant duration and number of data collection campaigns, coupled with the consequent different plant development stages, this dataset should provide an adequate surrogate for on-farm pasture conditions.

As spectral data is highly correlated, it is unlikely that the relationships found would not perform equally well whenever employed in perennial ryegrass pastures displaying the same range of biophysical and biochemical status to the sample site in this study, granted that no major factor which could influence spectral response is introduced (e.g. changing light/atmospheric conditions or soil-background at low biomass levels). This indicates a satisfactory approach to estimate a wide range of pasture biomass values throughout the year with a small number of VIs, which could be implemented on a UAV mounted sensor.

Conclusions

In this study, the accuracy and precision of vegetation indices and machine learning methods for biomass assessment were examined, while featuring the rising plate meter (RPM) as benchmark for pasture management. This study was able to present, test and evaluate an optimal method to select and employ spectral vegetation indices (VIs) for biomass assessment of perennial ryegrass under different fertilisation, regrowth periods and seasonal effects. Findings are in-line with previous results reported in the literature, while the best method largely outperforms standard practices (i.e. RPM) as well as known VIs.

The study has shown that the optimised NRI is equivalent to the RPM while providing a firm analytical ground for substitution between these techniques. Also, it is clear that reflectance can outperform canopy height as a feature for biomass estimation provided that the optimal spectral bands and methods are employed. Furthermore, multispectral systems can be employed in a multitude of data acquisition platforms, providing estimates at near or real-time.

This analysis could lead to the development of a multispectral UAV-mounted sensor with known added-value in relation to current on-farm practices and precision agriculture methods. Overall, the comparison of achievable accuracies by employing either canopy height or reflectance (through VIs) can indicate the most appropriated path when developing sensors and establishing optimal sensing techniques (either spectral or canopy based) for assessing pasture biomass.

Given the agreement of these findings with previous research, the best method (STACK and selected VIs, as feature(s)-algorithm pair) and validation results (i.e. assessed against an unseen validation set) should be transferable from this study site to other locations. Furthermore, from an end-user perspective, the accuracy and precision achieved is sufficient to efficient pasture-management.

Given the high accuracy achieved, further improvements to this methodology will necessarily have to take into account the accuracy of reference measurements and data-collection protocols. However, given the unsatisfactory results of current VI methods, a new UAV sensor (with a small number of bands) should benefit from the results and methods presented here. Finally, this study also indicated that the most well-known VI (NDVI) is not optimal for biomass estimation. It also indicated that better performing VIs which could be measured from a UAV-mounted multispectral sensor, which would allow mapping and monitoring of pasture biomass with high accuracy at very high spatial resolution and with complete coverage of management areas.

Acknowledgements This research was supported by Dairy Australia, through the Dairy on PAR action. All analysis was performed in *R 3.5.1*, mostly using the *hdsar* and *caret* packages, without which this research would not be possible.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not

permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aasen, H., Honkavaara, E., Lucieer, A., & Zarco-Tejada, P. (2018). Quantitative remote sensing at ultra-high resolution with UAV spectroscopy: A review of sensor technology, measurement procedures, and data correction workflows. *Remote Sensing*, *10*(7), 1091. <https://doi.org/10.3390/rs10071091>.
- Alckmin, G. T., Kooistra, L., Lucieer, A., & Rawnsley, R. (2019). Feature filtering and selection for dry matter estimation on perennial ryegrass: A case study of vegetation indices. In *Proceedings of the ISPRS-International Archives of The Photogrammetry, Remote Sensing and Spatial Information Sciences* (Vols. XLII-2/W13, pp. 1827–1831). <https://doi.org/10.5194/isprs-archives-XLII-2-W13-1827-2019>.
- Ali, I., Cawkwell, F., Dwyer, E., Barrett, B., & Green, S. (2016). Satellite remote sensing of grasslands: From observation to management. *Journal of Plant Ecology*, *9*(6), 649–671. <https://doi.org/10.1093/jpe/rtw005>.
- Allen, V. G., Batello, C., Berretta, E. J., Hodgson, J., Kothmann, M., Li, X., et al. (2011). An international terminology for grazing lands and grazing animals. *Grass and Forage Science*, *66*(1), 2–28. <https://doi.org/10.1111/j.1365-2494.2010.00780.x>.
- Breiman, L. (2001). *Random forests* (45th ed., Vol. 45, pp. 5–32). Hingham, MA, USA: Springer. <https://doi.org/10.1023/A:1010933404324>.
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (2017). *Classification and regression trees* (pp. 1–358). Boca Raton, FL, USA: CRC Press. <https://doi.org/10.1201/9781315139470>.
- Burkart, A., Cogliati, S., Schickling, A., & Rascher, U. (2014). A novel UAV-based ultra-light weight spectrometer for field spectroscopy. *IEEE Sensors Journal*, *14*(1), 62–67. <https://doi.org/10.1109/JSEN.2013.2279720>.
- Caruana, R., Niculescu-Mizil, A., Crew, G., & Ksikes, A. (2004). Ensemble selection from libraries of models. In *Proceedings of 21st International Conference on Machine Learning—ICML '04* (p. 18). New York, New York, USA: ACM Press. <https://doi.org/10.1145/1015330.1015432>.
- Chapman, D. F., Tharmaraj, J., Agnusdei, M., & Hill, J. (2012). Regrowth dynamics and grazing decision rules: Further analysis for dairy production systems based on perennial ryegrass (*Lolium perenne* L.) pastures. *Grass and Forage Science*, *67* (1), 77–95. <https://doi.org/10.1111/j.1365-2494.2011.00824.x>.
- Cho, M. A., Skidmore, A., Corsi, F., van Wieren, S. E., & Sobhan, I. (2007). Estimation of green grass/herb biomass from airborne hyperspectral imagery using spectral indices and partial least squares regression. *International Journal of Applied Earth Observation and Geoinformation*, *9*(4), 414–424. <https://doi.org/10.1016/j.jag.2007.02.001>.
- Christie, K. M., Smith, A. P., Rawnsley, R. P., Harrison, M. T., & Eckard, R. J. (2018). Simulated seasonal responses of grazed dairy pastures to nitrogen fertilizer in SE Australia: Pasture production. *Agricultural Systems*, *166*, 36–47. <https://doi.org/10.1016/j.agsy.2018.07.010>.
- Clevers, J. G. W., Heijden, G. W. M. van der, Verzakov, S., & Schaepman, M. E. (2007). Estimating grassland biomass using SVM band shaving of hyperspectral data. *Photogrammetric Engineering & Remote Sensing*, *73* (10), 1141–1148. <https://doi.org/10.14358/PERS.73.10.1141>.
- Collins, W. (1978). Remote sensing of crop type and maturity. *Photogrammetric Engineering and Remote Sensing*, *44*(1), 42–55.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. <https://doi.org/10.1007/BF00994018>.
- Darvishzadeh, R., Skidmore, A., Schlerf, M., Atzberger, C., Corsi, F., & Cho, M. (2008). LAI and chlorophyll estimation for a heterogeneous grassland using hyperspectral measurements. *ISPRS Journal of Photogrammetry and Remote Sensing*, *63*(4), 409–426. <https://doi.org/10.1016/j.isprsjprs.2008.01.001>.
- Delegido, J., Verrelst, J., Alonso, L., & Moreno, J. (2011). Evaluation of sentinel-2 red-edge bands for empirical estimation of green LAI and chlorophyll content. *Sensors*, *11*(7), 7063–7081. <https://doi.org/10.3390/s110707063>.
- Earle, D., & McGowan, A. (1979). Evaluation and calibration of an automated rising plate meter for estimating dry matter yield of pasture. *Australian Journal of Experimental Agriculture*, *19*(98), 337. <https://doi.org/10.1071/EA9790337>.
- Friedman, J. H. (1991). Multivariate adaptive regression splines. *The Annals of Statistics*, *19*(1), 1–67. <https://doi.org/10.1214/aos/1176347963>.
- García, S. C., Islam, M. R., Clark, C. E. F., & Martin, P. M. (2014). Kikuyu-based pasture for dairy production: A review. *Crop and Pasture Science*, *65*(8), 787. <https://doi.org/10.1071/CP13414>.

- Guyot, G., & Baret, F. (1988). Utilisation de la haute résolution spectrale pour suivre l'état des couverts végétaux. (Use of high spectral resolution for vegetation monitoring). In *Proceedings of the 4th International Colloquium on Spectral Signatures of Objects in Remote Sensing* (Vol. 287, pp. 279–286). D. Guyenne & J. J. Hunt (Eds.). New York, USA.
- Hall, A., Turner, L., & Kilpatrick, S. (2019). Understanding Tasmanian dairy farmer adoption of pasture management practices: A Theory of Planned Behaviour approach. *Animal Production Science*, 59(10), 1941–1950. <https://doi.org/10.1071/AN18321>.
- Horler, D. N. H., Dockray, M., & Barber, J. (1983). The red edge of plant leaf reflectance. *International Journal of Remote Sensing*, 4(2), 273–288. <https://doi.org/10.1080/01431168308948546>.
- Joint Committee for Guides in Metrology. (2008). Evaluation of measurement data—Guide to the expression of uncertainty in measurement. Retrieved 23 March 2020 from: <https://www.bipm.org/en/publications/guides/gum.html>.
- Kawamura, K., Betteridge, K., Costall, D., Sanches, I. D., Tuohy, M. P., & Inoue, Y. (2008). Spectro-CAPP as a tool to estimate and map pasture biomass and mineral component in New Zealand. In *Proceedings of the 29th Asian Conference on Remote Sensing 2008, ACRS, 2008* (Vol. 2, pp. 1345–1350). Tokyo, Japan: Asian Association on Remote Sensing
- Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, 28(5), 1–26. <https://doi.org/10.18637/jss.v028.i05>.
- Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling* (Vol. 26, pp. 1–600). New York, USA: Springer. <https://doi.org/10.1007/978-1-4614-6849-3>.
- Lehnert, L. W., Meyer, H., Obermeier, W. A., Silva, B., Regeling, B., & Bendix, J. (2018). Hyperspectral data analysis in R: the hsdar Package. *Journal of Statistical Software*. <https://doi.org/10.18637/jss.v089.i12>.
- Lichtenthaler, H. K., Lang, M., Sowinska, M., Heisel, F., & Miehe, J. A., (1996). Detection of vegetation stress via a new high resolution fluorescence imaging system. *Journal of Plant Physiology*, 148(5), 599–612. [https://doi.org/10.1016/S0176-1617\(96\)80081-2](https://doi.org/10.1016/S0176-1617(96)80081-2).
- Maire, G., François, C., Soudani, K., Berveiller, D., Pontailier, J. Y., Bréda, N., et al. (2008). Calibration and validation of hyperspectral indices for the estimation of broadleaved forest leaf chlorophyll content, leaf mass per area, leaf area index and leaf canopy biomass. *Remote Sensing of Environment*, 112(10), 3846–3864. <https://doi.org/10.1016/j.rse.2008.06.005>.
- Massey, F. J. (1951). The Kolmogorov–Smirnov test for goodness of fit. *Journal of the American Statistical Association*, 46(253), 68–78. <https://doi.org/10.1080/01621459.1951.10500769>.
- Mayer, Z. A., & Knowles, J. E. (2015). *caretEnsemble: Ensembles of caret models*. Retrieved 23 March 2020 from <https://cran.r-project.org/web/packages/caretEnsemble/index.html>.
- Michez, A., Lejeune, P., Bauwens, S., Herinaina, A., Blaise, Y., Castro Muñoz, E., et al. (2019). Mapping and monitoring of biomass and grazing in pasture with an unmanned aerial system. *Remote Sensing*, 11(5), 473. <https://doi.org/10.3390/rs11050473>.
- Mokvist, F., Mamedov, F., & Styring, S. (2014). Defining the Far-red limit of photosystem I: The primary charge separation is functional to 840 nm. *Journal of Biological Chemistry*, 289(35), 24630–24639. <https://doi.org/10.1074/jbc.M114.555649>.
- Mutanga, O., & Skidmore, A. K. (2004). Narrow band vegetation indices overcome the saturation problem in biomass estimation. *International Journal of Remote Sensing*, 25(19), 3999–4014. <https://doi.org/10.1080/01431160310001654923>.
- Nakagami, K. (2016). Effects of sites and years on the coefficients of rising plate meter calibration under varying coefficient models. *Grassland Science*, 62(2), 128–132. <https://doi.org/10.1111/grs.12117>.
- Perez-Riverol, Y., Kuhn, M., Vizcaíno, J. A., Hitz, M. P., & Audain, E. (2017). Accurate and fast feature selection workflow for high-dimensional omics data. *PLoS ONE*, 12(12), 1–14. <https://doi.org/10.1371/journal.pone.0189875>.
- Thenkabail, P. S., Smith, R. B., & De Pauw, E. (2002). Evaluation of narrowband and broadband vegetation indices for determining optimal hyperspectral wavebands for agricultural crop characterization. *Photogrammetric Engineering and Remote Sensing*, 68(6), 607–621.
- Thomson, N. A., Mccallum, D. A., Howse, S., Holmes, C. W., Matthews, P. N. P., & Matthew, C. (1997). Estimation of dairy pastures - the need for standardisation. In *Proceedings of the New Zealand Grassland Association* (Vol. 59, pp. 221–225).
- Tucker, C. J. (1977). Spectral estimation of grass canopy variables. *Remote Sensing of Environment*, 6(1), 11–26. [https://doi.org/10.1016/0034-4257\(77\)90016-5](https://doi.org/10.1016/0034-4257(77)90016-5).
- Turner, L. R., Donaghy, D. J., Lane, P. A., & Rawnsley, R. P. (2006). Effect of defoliation management, based on leaf stage, on perennial ryegrass (*Lolium perenne* L.), prairie grass (*Bromus willdenowii* Kunth.) and cocksfoot (*Dactylis glomerata* L.) under dryland conditions: Nutritive value. *Grass and Forage Science*, 61(2), 175–181. <https://doi.org/10.1111/j.1365-2494.2006.00524.x>.

- Wang, S., Baum, A., Zarco-Tejada, P. J., Dam-Hansen, C., Thorseth, A., Bauer-Gottwein, P., et al. (2019). Unmanned Aerial System multispectral mapping for low and variable solar irradiance conditions: Potential of tensor decomposition. *ISPRS Journal of Photogrammetry and Remote Sensing*, *155*, 58–71. <https://doi.org/10.1016/j.isprsjprs.2019.06.017>.
- Xue, J., & Su, B. (2017). Significant remote sensing vegetation indices: A review of developments and applications. *Journal of Sensors*, *2017*, 1–17. <https://doi.org/10.1155/2017/1353691>.
- Zarco-Tejada, P. J., Pushnik, J. C., Dobrowski, S., & Ustin, S. L. (2003). Steady-state chlorophyll a fluorescence detection from canopy derivative reflectance and double-peak red-edge effects. *Remote Sensing of Environment*, *84*(2), 283–294. [https://doi.org/10.1016/S0034-4257\(02\)00113-X](https://doi.org/10.1016/S0034-4257(02)00113-X).

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.