

## Article

# UAV remote sensing for high-throughput phenotyping and for yield prediction of *Miscanthus* by machine learning techniques

Giorgio Impollonia<sup>1,2\*</sup>, Michele Croci<sup>1,2</sup>, Andrea Ferrarini<sup>1</sup>, Jason Brook<sup>3</sup>, Enrico Martani<sup>1</sup>, Henri Blandinières<sup>1</sup>, Andrea Marcone<sup>1,2</sup>, Danny Awty-Carroll<sup>4</sup>, Chris Ashman<sup>4</sup>, Jason Kam<sup>5</sup>, Andreas Kiesel<sup>6</sup>, Luisa M. Trindade<sup>7</sup>, Mirco Boschetti<sup>8</sup>, John Clifton-Brown<sup>9</sup>, Stefano Amaducci<sup>1,2</sup>

- <sup>1</sup> Department of Sustainable Crop Production, Università Cattolica del Sacro Cuore, Piacenza 29122, Italy; [andrea.ferrarini@unicatt.it](mailto:andrea.ferrarini@unicatt.it); [enrico.martani@unicatt.it](mailto:enrico.martani@unicatt.it); [henri.blandinieres@unicatt.it](mailto:henri.blandinieres@unicatt.it); [stefano.amaducci@unicatt.it](mailto:stefano.amaducci@unicatt.it)
- <sup>2</sup> Remote Sensing and Spatial Analysis Research Center (CRAS), Università Cattolica del Sacro Cuore, Piacenza 29122, Italy; [giorgio.impollonia@unicatt.it](mailto:giorgio.impollonia@unicatt.it); [michele.croci@unicatt.it](mailto:michele.croci@unicatt.it); [andrea.marcone@unicatt.it](mailto:andrea.marcone@unicatt.it)
- <sup>3</sup> The National Plant Phenomics Centre, Institute of Biological, Environmental and Rural Sciences (IBERS), Aberystwyth University, Aberystwyth, United Kingdom; [jkb@aber.ac.uk](mailto:jkb@aber.ac.uk)
- <sup>4</sup> Aberystwyth University, Institute of Biological, Environmental and Rural Sciences, Aberystwyth, SY23 3EE, UK; [dga1@aber.ac.uk](mailto:dga1@aber.ac.uk); [cha4@aber.ac.uk](mailto:cha4@aber.ac.uk)
- <sup>5</sup> Terravesta, Unit 4 Riverside Court, Skellingthorpe Road, Saxilby, Lincoln LN1 5AB; [jason@terravesta.com](mailto:jason@terravesta.com)
- <sup>6</sup> Department of Biobased Resources in the Bioeconomy, Institute of Crop Science, University of Hohenheim, Stuttgart, Germany; [a.kiesel@uni-hohenheim.de](mailto:a.kiesel@uni-hohenheim.de)
- <sup>7</sup> Wageningen University & Research, Plant Breeding, 6700 AJ Wageningen, The Netherlands; [luisa.trindade@wur.nl](mailto:luisa.trindade@wur.nl)
- <sup>8</sup> Institute for Electromagnetic Sensing of the Environment, National Research Council, Milan, Italy; [boschetti.m@irea.cnr.it](mailto:boschetti.m@irea.cnr.it)
- <sup>9</sup> Justus-Liebig-Universität Gießen. iFZ, Heinrich-Buff-Ring 26, 35392 Gießen, Germany; [john.clifton-brown@agr.uni-giessen.de](mailto:john.clifton-brown@agr.uni-giessen.de)

\* Correspondence: [giorgio.impollonia@unicatt.it](mailto:giorgio.impollonia@unicatt.it)

**Abstract:** *Miscanthus* holds a great potential in the frame of the bioeconomy and yield prediction can help improving *Miscanthus* logistic supply chain. Breeding programs in several countries are attempting to produce high-yielding *Miscanthus* hybrids better adapted to different climates and end-uses. Multispectral images acquired from unmanned aerial vehicles (UAVs) in Italy and in the UK in 2021 and 2022 were used to investigate the feasibility of high-throughput phenotyping (HTP) of novel *Miscanthus* hybrids for yield prediction and crop traits estimation. An intercalibration procedure was performed using simulated data from the PROSAIL model to link vegetation indices (VIs) derived from two different multispectral sensors. Random forest algorithm estimated with good accuracy yield traits (light interception, plant height, green leaf biomass and standing biomass) using VIs time series and predicted yield using peak descriptor derived from VIs time series with 2.3 Mg DM ha<sup>-1</sup> of RMSE. The study demonstrates the potential of UAVs multispectral in HTP applications and in yield prediction for providing important information needed to increase sustainable biomass production.

**Keywords:** *Miscanthus*; remote sensing; UAV; multispectral images; high-throughput phenotyping; machine learning; yield prediction; trait estimation; PROSAIL; multi-sensor interoperability

## 1. Introduction

*Miscanthus* is a high yielding perennial biomass crop. Yield is one of the most important traits of *Miscanthus* [1] and has been the primary focus of the research portfolio on *Miscanthus* in the last ten years [2–4]. Yield depends not only on the climatic and soil characteristics but also on crop age, harvest date and the genotype/hybrid type [4]. The rate of mature yield development varies with climate, but in general increases from 1-3 Mg ha<sup>-1</sup> of dry matter (DM) in the first year to 6-8 Mg DM ha<sup>-1</sup> in the second year and to 12-30 Mg

DM ha<sup>-1</sup> in the third year and onwards in rainfed Northern and rainfed or irrigated Southern Europe [3,5–7]. Independent and collaborative efforts to breed high-yielding *Miscanthus* hybrids to produce sustainable biomass for the growing bio-based European market are ongoing in several countries [8–10]. In the EU-BBI demo-project GRACE<sub>2</sub> novel highly upscalable seed-based *Miscanthus* hybrids are evaluated [8,10–13] in seven European countries. Most yield prediction to date has relied on crop growth models driven by climate and soil data with crop specific parameters [14–16]. For example, MISCANFOR is a crop growth model specifically developed to predict *Miscanthus x giganteus* yields in a wide range of environments [17]. It has been widely used and validated at European [9] and national level [18] for *Miscanthus* and other perennial biomass crops [19], but new parameterization data to predict yield production of the novel *Miscanthus* hybrids [8] is required [17].

Yield trait screening and prediction using remote sensing with unmanned aerial vehicles (UAVs) can help both in breeding activities and in obtaining spatial and temporal information for optimising *Miscanthus* biomass supply chain logistics, from field to facilities creating bioproducts or biopower [20,21]. Impollonia *et al.* [22] recently demonstrated the feasibility of moisture content estimation in *Miscanthus* hybrids using vegetation indices (VIs) derived from UAV-based remote sensing. Remote sensing approaches can also be used to i) estimate yield-related traits for high-throughput phenotyping (HTP) [23–25], ii) calibrate crop growth models [26,27] and iii) predict the yield of many crops for commercial purposes [28]. To date, few studies have focused on the estimation and prediction of perennial biomass crops traits using remote sensing technologies from satellite or UAV [29–31] and only two on *Miscanthus* [21,22]. Crop traits such as the plant height [32], the fraction of absorbed photosynthetically active radiation (fAPAR) [33] and the above-ground biomass [34] can be estimated from the VIs in combination with machine learning (ML) algorithms. One of the most used ML algorithms in remote sensing analyses for crop traits estimation is random forest (RF) [35–37]. RF proved to be robust to outliers and noise, does not suffer from overfitting, and can manage a high training size [38]. The use of ML algorithms shows great potentials in crop yield prediction [39–42]. In particular, the use of VIs' time series helped to derive descriptors of land surface phenology (LSP, i.e. the spatial and temporal development of the vegetated land surface) [43–45] such as the start of season (SOS), the peak of growing season, the stay-green duration (onset of senescence), the end of the season (EOS), and growing season length [46]. Among the available descriptors of LSP, the peak of a VI is one of the most important descriptor for crop yield prediction, such as the peak of NDVI [47] and EVI2 [20] for grain yield and the peak of GNDVI for biomass yields of perennial grass [29]. Yield prediction is, in many studies, theoretical, as it requires to fit the whole seasonal curve for deriving the peak of the time series of VIs. This renders impossible the yield prediction before having obtained the data of the whole seasonal time series. Being able to perform yield prediction months before harvest with partial time series should permit to optimize *Miscanthus* biomass supply chain logistics for practical applications.

In addition, the VIs values used for estimation and prediction of crop traits are influenced by many factors, such as sensor characteristics, atmospheric conditions during acquisition, viewing angle, field of view, and sun elevation [48]. In the context of crop phenotyping, where the field trials are often carried out in multi-location and with different UAV sensors, these factors could have a relevant effect on the compatibility of VIs. Among different sensors characteristics, the full width at half maximum (FWHM) is the main factor that influences the comparability of VIs values among different sensors [49]. Indeed, due to the different spectral characteristics of the UAV multispectral sensors available on the market, differences among VIs derived from multiple UAV sensors for the same target can be found [48]. For this reason, there is a need to increase the interoperability of the sensors using equations able to overcome these differences through advanced linking procedures between the VIs of sensors [50,51]. The multi-sensor interoperability is an important topic in remote sensing science [52–54] when multi-location monitoring is conducted. Overcoming this problematic can be realised by intercalibrating the VIs obtained

by different sensors and evaluating the correlation through linear regressions [55–57]. This intercalibration can be realised in two main ways: a direct approach where VIs are measured by sensors [58] and an indirect approach where VIs data are simulated by radiative transfer models [59]. A major limitation of the direct regression approach is that it is not transferable because it is site-specific [50]. On the contrary, the indirect approach permits to retrieve VIs data from radiative transfer models such as the PROSAIL model [60,61]. PROSAIL has been used to assess the performances of different satellite sensors for multiple VIs [62] but a similar application is currently lacking in UAV science. Hence, the use of such approach for VIs intercalibration in UAV science might allow to overcome the problematic of multi-sensor interoperability across different sites.

In summary, the overall objectives of this study based on UAV remote sensing were: i) to estimate crop traits (light interception, plant height, green leaf biomass and standing biomass) for supporting breeding programs and for providing modelling parameters for *Miscanthus*, ii) to predict yield to obtain spatial and temporal information for improving the logistics biomass supply chain of *Miscanthus*, and iii) to explore the potential impact of the timeliness on the yield prediction, by evaluating the performance of the yield prediction model using peak derived from partial time series of VIs. To achieve these overarching objectives, UAV multispectral images and ground phenotypic data were collected at two locations within the same multi-environment trial: one in Italy and one in the UK. These data were analysed using: i) the PROSAIL model to simulate crop spectral signatures in order to intercalibrate VIs of two different common multispectral sensors, ii) random forest (RF) algorithm to estimate crop traits using the VIs time series and to predict yield using the peak descriptor derived from the VIs time series, for the reliability it achieved in previous studies [20,29,47], and iii) generalized additive model (GAM) to derive peak from complete and partial time series of VIs.

## 2. Materials and Methods

### 2.1. Experimental design

The field trials were conducted in two locations: PAC 1 located in San Bonico in the Italian province of Piacenza (NW Italy) (45°00'11.70" N, 9°42'35.39" E) and TWS 1 located in Trawscoed near Aberystwyth in Wales (UK) (52°24'59.8"N, 4°04'02.6"W). These sites are two of the seven plot scale (PS) trials conducted within EU-BBI GRACE demo-project. In PAC 1 the climate is temperate with a mean annual precipitation of 792 mm, while the climate in TWS 1 is oceanic with a mean annual precipitation of 984 mm. The trials were established in April 2018 with 14 *Miscanthus* hybrids with  $n = 4$  replicates for a total of  $n = 56$  plots. The trials were planted with eight novel intraspecific *M. sinensis*  $\times$  *M. sinensis* hybrids (*M. sin*  $\times$  *M. sin*), five novel interspecific hybrids *M. sinensis*  $\times$  *M. sacchariflorus* (*M. sin*  $\times$  *M. sac*) and *M. x giganteus* as control genotype (for more details see Impollonia *et al.*[22]).

### 2.2. Phenotypic and yield measurements

Phenotypic measurements were taken from the emergence of the crop in spring 2020 until the winter harvest in the early months of 2021. This season will thereafter be referred to as the 2020 growing season. These phenotypic measurements were carried out in the two locations and on four out of the fourteen hybrids in the trial: GRC 3 (*M. sin*  $\times$  *M. sin*), GRC 14 (seeded) and GRC 15 (clonal) (*M. sin*  $\times$  *M. sac*) and GRC 9 (standard clonal *M. x giganteus*). The measurements of this study were carried during 3<sup>rd</sup> growing season. Five contiguous plants along a central row in each plot were used for monitoring along the growing season. The following list of crop traits were measured along the growing season: plant height (cm) and light interception (%) were measured weekly; green leaf biomass (Mg DM ha<sup>-1</sup>) and standing biomass (Mg DM ha<sup>-1</sup>) were measured fortnightly. One hundred seventy-two and 145 data were collected for light interception, 240 and 204 for plant height, 232 and 316 for green leaf biomass, and 268 and 328 for standing biomass, in PAC 1 and TWS 1 respectively. Plant height was measured from the soil to the height of the last ligule of the tallest stem using a graduated pole until crop reached complete flowering or started to senesce in November. Light interception was measured with a lab-constructed

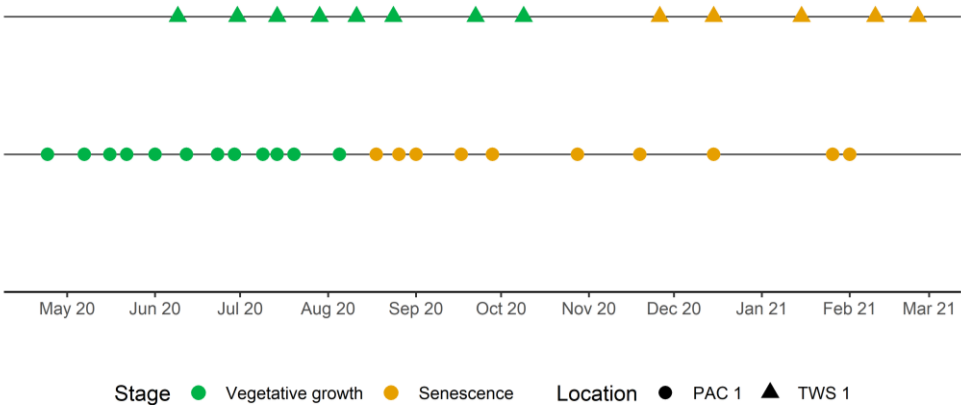
1m “line ceptometer” with 10 photodiodes at 10 cm spacings generating an electric current which is converted with simple circuitry to a voltage linearly proportional to the light intensity. Light intensity was measured above the canopy and at the base of each of the five selected plants. Light interception measurements were carried out from emergence until full canopy closure (around 95% of light is intercepted by the crop canopy) on a weekly basis and later at a lower frequency. Standing aboveground biomass was estimated on a monthly basis, starting after emergence in 2020 until harvest in winter 2021 using the methods described in Magenau *et al.* [63]. In brief, 10 randomly sampled shoots per plot each fortnight (aka ‘serial cuts’) were related to the final quadrat yields at spring harvest and used to back calculate the seasonal dynamics of above ground biomass (Mg DM ha<sup>-1</sup>) from spring emergence until final harvest in the following spring. Each 10 shoot serial cuts were separated into green leaf, brown leaf, and stem fractions. The fresh weight measured in field, and the dry weight measured after oven drying to constant weight at 80°C were used to calculate the mass and moisture contents of each fraction and were scaled to Mg DM ha<sup>-1</sup>. The crop phenological stages were estimated using the thermal time following the method proposed on *M. x giganteus* by Tejera *et al.* [64]. Thermal time was measured in growing degree days (GDD, °C Day) as:

$$GDD = \left[ \frac{T_{max} + T_{min}}{2} \right] - T_b \quad (1)$$

where  $T_{max}$  and  $T_{min}$  are the daily highest and lowest temperatures measured by a weather station situated at each experimental site, and  $T_b$  is a base growth temperature of 6°C [65]. Two main phenological stages were estimated: vegetative growth and senescence. The GDD accumulation started at plant emergence, and the accumulation of 1500 GDD was used as the threshold marking the difference between the vegetative growth and senescence. The final harvested yield was measured for all the fourteen hybrids using a quadrat area of 6.6 m<sup>2</sup> (10 plants per plot planted at 1.5 plants m<sup>-2</sup> (*M. x giganteus* and *M. sin x M. sac*) or 20 plants per plots at 3 plants m<sup>-2</sup> (*M. sin x M. sin*)) and a cutting height of 10 cm. In each plot, the fresh weight of all plants in the quadrat was recorded and a sub-sample of approximately five stems per plot was used to determine the moisture content and calculate the yield in Mg DM ha<sup>-1</sup>. Plants were harvested on February 2<sup>nd</sup> 2021 (Days of year (DOY): 33) at PAC 1 and on March 8<sup>th</sup> 2021 (DOY: 67) at TWS 1.

### 2.3. UAV multispectral data and vegetation indices

Unmanned aerial vehicle (UAV) multispectral data acquisition flights were performed approximately fortnightly from 24/04/2020 to 01/02/2021 at the PAC 1 (25 flights) site and from 09/06/2020 to 25/02/2021 at the TWS 1 (17 flights) site (Figure 1). Table 1 details the specifications of the UAVs and the multispectral cameras used at the two sites. All the flights were performed between 11 am and 3 pm with flight altitude above ground level (AGL) fixed at 50 m and 40 m at PAC 1 and TWS 1, respectively. The forward and lateral overlap was set at 80% and 75% of the images, respectively. Spectral panels and light sensors mounted on top of the UAVs were used for the radiometric calibration of the images. The radiometric calibration and orthomosaic generation were done using the Pix4D mapper (Pix4D, S.A., Lausanne, Switzerland). The 15 vegetation indices (VIs) were calculated as shown in Table 2 using the orthomosaic. The mean of the VIs was extracted for each plot using the polygons of the experimental designs that were drafted in AutoCAD (Autodesk, San Rafael, California, USA) and georeferenced with QGIS software (QGIS Development Team, 2021).



**Figure 1** Seasonal calendar of UAV flights performed in the two locations according to the estimated phenological stages (vegetative growth and senescence).

**Table 1** Unmanned aerial vehicles (UAVs) and multispectral cameras used to perform flights in the two locations, with their respective characteristics of spectral bands, central wavelength, and full width at half maximum (FWHM).

Location	UAV	Multispectral camera characteristics			
		Model	Band	Centre (nm)	FWHM (nm)
PAC 1	DJI M210 RTK	MicaSense RedEdge-Mx	Blue	475	32
			Green	560	27
			Red	668	14
			Red edge	717	12
			Near-infrared	840	57
TWS 1	DJI M210	SlantRange 4P	Blue	470	100
			Green	550	100
			Red	650	40
			Red edge	710	20
			Near-infrared	850	100

**Table 2** List of the vegetation indices evaluated in random forest models for crop traits estimation and for yield prediction of *Miscanthus*.

VIs	Equation	Reference
Datt1	$\frac{\text{NIR} - \text{RedEdge}}{\text{NIR} + \text{Red}}$	[66]
EVI2	$2.4 \frac{\text{NIR} - \text{Red}}{1 + \text{NIR} + \text{Red}}$	[67]
GNDVI	$\frac{\text{NIR} - \text{Green}}{\text{NIR} + \text{Green}}$	[68]
GOSAVI	$\frac{\text{NIR} - \text{Green}}{\text{NIR} + \text{Green} + 0.16}$	[69]
greenWDRVI	$\frac{0.1\text{NIR} - \text{Green}}{0.1\text{NIR} + \text{Green}} + \frac{1 - 0.1}{1 + 0.1}$	[70]
MSAVI	$\frac{2\text{NIR} + 1 - \sqrt{(2\text{NIR} + 1)^2 - 8(\text{NIR} - \text{Red})}}{2}$	[71]
MTVI1	$1.2(1.2(\text{NIR} - \text{Green}) - 2.5(\text{Red} - \text{Green}))$	[72]
MTVI2	$1.5 \frac{1.2(\text{NIR} - \text{Green}) - 2.5(\text{Red} - \text{Green})}{\sqrt{(2\text{NIR} + 1)^2 - 6\text{NIR} - 5\sqrt{\text{Red}} - 0.5}}$	[72]
NDRE	$\frac{\text{NIR} - \text{RedEdge}}{\text{NIR} + \text{RedEdge}}$	[73]
NDVI	$\frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red}}$	[74]
OSAVI	$(1 + 0.16) \frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red} + 0.16}$	[75]
OSAVI2	$(1 + 0.16) \frac{\text{NIR} - \text{RedEdge}}{\text{NIR} + \text{RedEdge} + 0.16}$	[76]
rededgeWDRVI	$\frac{0.1\text{NIR} - \text{RedEdge}}{0.1\text{NIR} + \text{RedEdge}} + \frac{1 - 0.1}{1 + 0.1}$	[70]
SAVI	$(1 + 0.5) \frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red} + 0.5}$	[77]
WDRVI	$\frac{0.1\text{NIR} - \text{Red}}{0.1\text{NIR} + \text{Red}} + \frac{1 - 0.1}{1 + 0.1}$	[70]

#### 2.4. Using the PROSAIL model to intercalibrate VIs from different multispectral sensors

The PROSAIL model was used to simulate crop spectral signatures to intercalibrate VIs (Table 2) calculated from the two different multispectral sensors used in this study (Table 1). The PROSAIL can simulate the canopy reflectance, between 400 and 2500 nm, by combining the PROSPECT and SAIL models. The PROSPECT model simulates the optical properties of the leaves using 4 input parameters: leaf structure parameter (N), leaf chlorophyll content (LCC), relative leaf equivalent water thickness ( $C_{wr}$ ) and leaf dry matter content ( $C_m$ ). The SAIL model simulates the bidirectional reflectance of a canopy using 6 input parameters: leaf area index (LAI), leaf inclination distribution (LIDF), hotspot parameter (hot), solar zenith angle (tts), observer zenith angle (tto) and relative azimuth angle (psi). Canopy and leaf parameters for *Miscanthus* were retrieved from available data on literature [78,79]. The *hsdar* R package [80] was used to simulate the canopy reflectance of the PROSAIL model using the function *PROSAIL* which uses the FORTRAN code of the PROSAIL model (Version 5B). The look-up table (LUT) generated included 430,080 parameter combinations following the min-max ranges of input parameters summarized in Table 3.



**Table 3** Ranges of input parameters for the PROSAIL model for the generation of the LUT.

	Parameter	Abbreviation	Unit	Values (step)
Leaf	Structure parameter	N	Unitless	1 – 2 (1)
	Chlorophyll content	LCC	$\mu\text{g cm}^{-2}$	10 – 80 (10)
	Relative equivalent water thickness	$C_{wr}$	%	20 – 80 (20)
	Dry matter content	$C_m$	$\text{g cm}^{-2}$	0.01 – 0.025 (0.005)
Canopy	Leaf area index	LAI	$\text{m}^2 \text{m}^{-2}$	1 – 8 (1)
	Leaf inclination distribution	LIDF		Spherical
	Hotspot parameter	hot	$\text{m m}^{-1}$	0.05 – 0.45 (0.2)
	Solar zenith angle	tts	deg	20 – 80 (10)
	Observer zenith angle	tto	deg	5 – 10 (5)
	Relative azimuth angle	psi	deg	180 – 220 (10)
	Structure parameter	N	Unitless	1 – 2 (1)

The present work considers all potential parameters combinations for LUT generation because the novel *Miscanthus* hybrids evaluated have not previously been studied in literature and because the crop monitoring was performed on the whole growing season (vegetative growth and senescence). The spectral reflectance simulated were resampled based on UAV sensors characteristics (Table 1) and the 15 VIs used in this study were calculated. For each VI, regression analysis was conducted to intercalibrate the VI values calculated from the two multispectral sensors. Linear regressions were performed using the VIs and not the spectral bands, in order to i) evaluate the different sensitivity of the VIs to sensor characteristics and ii) identify which VIs need an intercalibration procedure between sensors. The final database of VIs was built by scaling SlantRange sensor VIs data toward MicaSense one.

### 2.5. Time series of VIs and peak derivation

The 15 VIs calculated from UAV multispectral images of the two sites and intercalibrated using the linear models derived by the PROSAIL simulation were smoothed using a generalized additive model (GAM) to generate daily VIs time series. The GAM is a non-parametric regression model which allows non-linear fitting of the variables. GAM models were fitted in R package “mgcv” [81]. GAM fittings of VIs allows to remove the outliers and regularize the time series [22,82,83]. The time series of VIs for each plot were fitted against the modified days of the year ( $\text{DOY}_M$ ). The  $\text{DOY}_M$  was used to overcome the problem of having non-sequential DOY throughout the growing season, as this last overlaps two different years.  $\text{DOY}_M$  were calculated as conventional DOY for the first year of the growing season (2020) and as  $\text{DOY} + 365$  starting from the 1<sup>st</sup> of January for the year 2021.

Daily time series of the VIs were used to estimate the crop traits by linking the traits values measured in the field with the VIs values of the time series. To predict yield, the peak descriptor was chosen among several land surface phenology (LSP) descriptors, due to the reliability it achieved in previous studies [20,29,47]. The peak descriptor is defined as the maximum value and was derived from the GAM fitting of each VI time series. Two types of VIs time series were evaluated to derive the peak descriptor value: complete and partial times series. Complete VIs time series were obtained by fitting the GAM models to the VIs data acquired throughout the entire growing season. Partial time series were obtained in two steps. Firstly, by fitting the GAM models to the VIs values obtained from a reduced set of initial UAV flights over the crop. The first seven flights (175 DOY in PAC 1 and 266 DOY in TWS 1) were selected to form this reduced set of initial UAV flights, based on physiology and phenology of *Miscanthus*, to cover most of the vegetative phase, the seventh flight being close to the peak of biomass accumulation. Secondly, the VIs data obtained from the following UAV flights were added one by one to the data fitted in the first step, up to the end of the season. The peak derived from complete time series of VIs were used for the yield prediction modelling and those from partial time series of VIs were used to analyse the variation of peak values and the model performance in order to assess feasibility of “early season” yield prediction during *Miscanthus* crop grow.

The random forest (RF) algorithm was used to estimate the crop traits (section 2.2) and to predict the yield of *Miscanthus* hybrids. The RF models were created using the *caret* R package [84]. Three steps in RF modelling were adopted: firstly, RF models were created for i) traits estimation using the 15 VIs of the 4 *Miscanthus* hybrids (see section 2.2) and phenological stages (encoded as 0 and 1 for the stages of vegetative growth and senescence, respectively) and ii) yield prediction using the peak values derived from complete time series of the 15 investigated VIs of the 14 *Miscanthus* hybrids; secondly, the importance of variables for each tested RF models was calculated by dropout loss of RMSE and the variables with a median of RMSE loss greater than 0 were selected; thirdly, the RF models used to estimate the variable of interest (i.e. crop traits or yield) were created using only the selected variables. The optimal size of the variable subset ("*mtry*") for each model was obtained by grid-searching method using repeated k-fold cross-validation (ten-fold cross validation repeated 5 times). The training dataset was created using a stratified random sampling method [34]: data from both locations and genotypes were split into 2/3 of the dataset for training and 1/3 for testing based on data distribution. The variable importance was calculated by dropout loss of RMSE (i.e. increase of prediction RMSE [85]) using the DALEX package [86]. To show the uncertainty of importance estimation, the variable importance was calculated for 10 permutations [85]. The RF models' performances were evaluated through the root mean square error (RMSE) and the normalized root mean square error (NRMSE), calculated as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n}} \quad (2)$$

$$NRMSE (\%) = \frac{\sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n}}}{\bar{y}} 100 \quad (3)$$

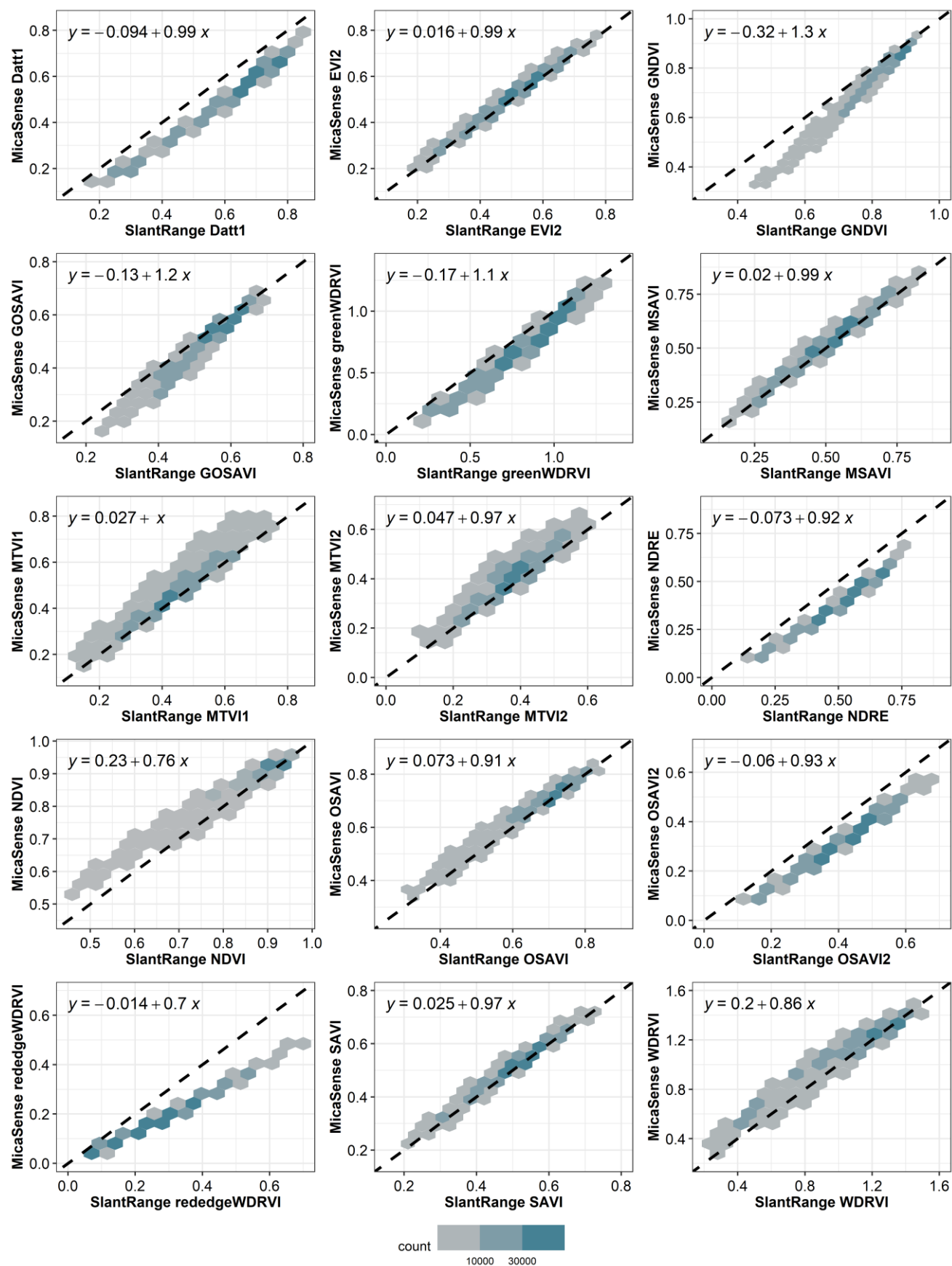
where  $n$  is the number of samples,  $x_i$  and  $y_i$  are the estimated and measured value of each trait, and  $\bar{y}$  is the mean of the measured values. RMSE and NRMSE metrics were also used to compare the performance of the yield prediction model using test datasets with the peak derived from complete and partial time series of VIs.

### 3. Results

#### 3.1. PROSAIL model for intercalibration of VIs derived from different multispectral sensors

The outputs of the linear regressions analysis for each VI of interest (see Table 2) as calculated from PROSAIL model simulation considering the specific spectral response function of the two sensors ( $y$  - the MicaSense VIs and  $x$  - the SlantRange VIs) are reported in Figure 2. EVI2, MSAVI and SAVI were the three VIs with the slope values closest to 1 and intercept values closest to 0 (Figure 2). The slope values of these VIs were 0.99 for EVI2, 0.99 for MSAVI and 0.97 for SAVI, while the intercept was respectively 0.016 for EVI2, 0.02 for MSAVI and 0.025 for SAVI. OSAVI had a similar relation (slope: 0.91) but showed higher variability at VI values lower than 0.6. Datt1, NDRE and OSAVI2 had a relationship between the two sensors with a slope close to 1 but they showed a different intercept indicating an offset of SlantRange sensor due to underestimation of the VI compared to MicaSense (Figure 2). A slope close to 1 associated with a high variability for the whole range of VI values was observed for MTVI1 (slope: 1) and MTVI2 (0.97). GNDVI, GOSAVI, greenWDRVI and NDVI showed instead the highest differences between the two sensors at the lowest values of VI. On the contrary, rededgeWDRVI showed the highest differences at high VI values.

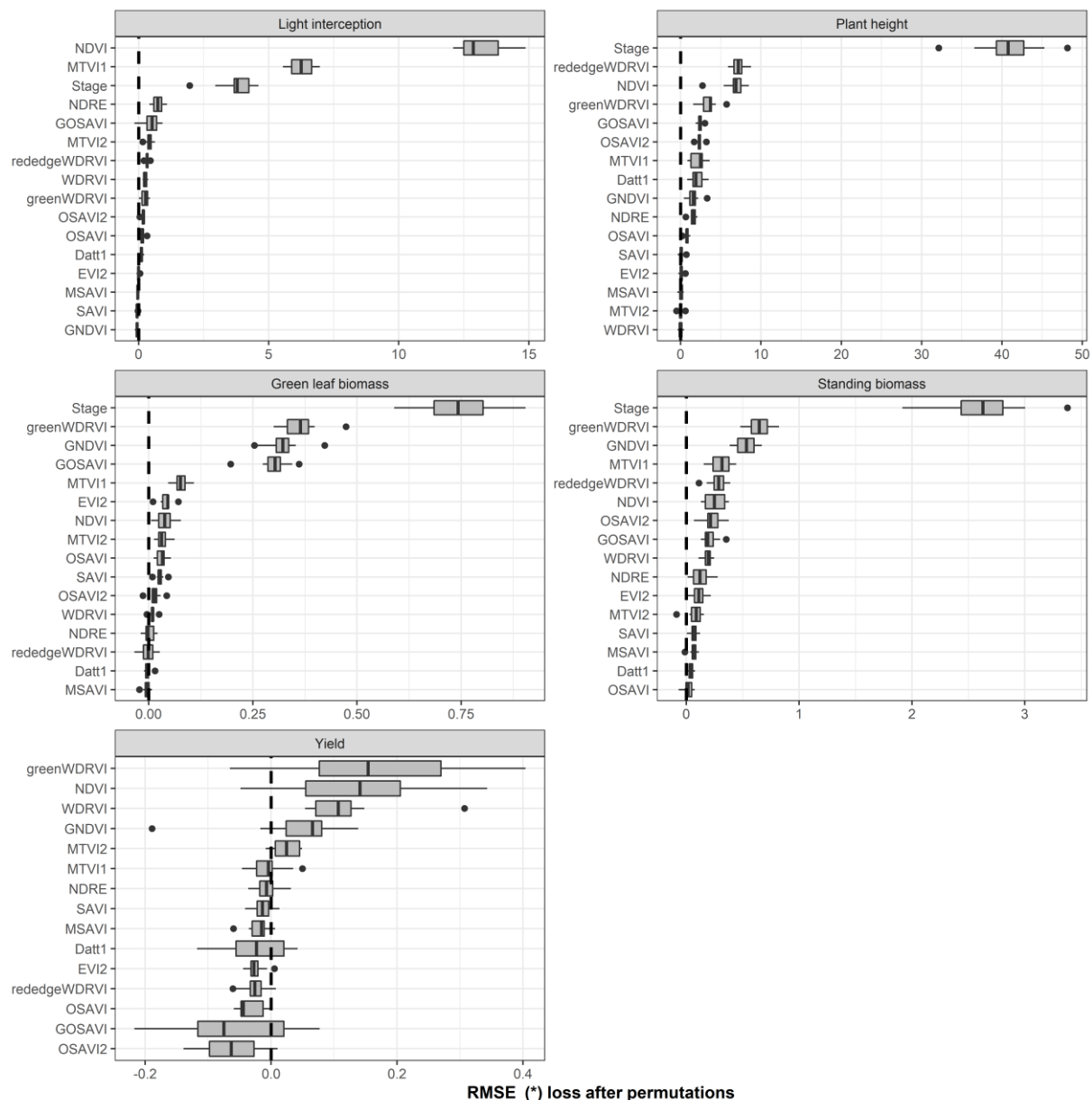




**Figure 2** Linear regression between VIs MicaSense and SlantRange derived from the PROSAIL model. The 1:1 relationship is represented by a dashed line.

### 3.2. Importance of variables in machine learning models

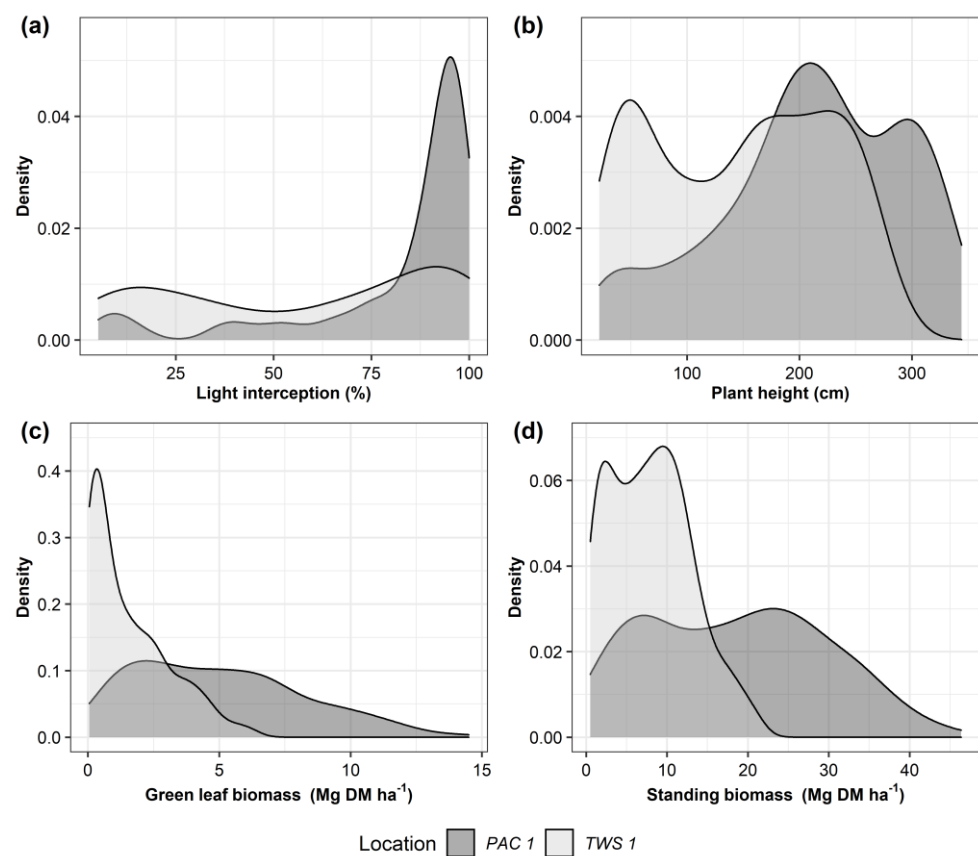
Importance of random forest (RF) models' variables for crop trait estimations and for yield prediction are shown in Figure 3. The analysis of variables' importance for the RF models was evaluated by drop-out loss of RMSE for each variable compared to the full model [85,86]. Phenological stage ("Stage") was the most important variable for estimating plant height, green leaf biomass and standing biomass (Figure 3). For crop trait estimations, the two most important VIs were NDVI and MTVI1 for light interception, rededgeWDRVI and NDVI for plant height, and greenWDRVI and GNDVI for green leaf biomass and standing biomass. For yield prediction using VIs peaks values, the most important VIs (with a median of RMSE loss greater than 0) were greenWDRVI, NDVI, WDRVI, GNDVI and MTVI2 (Figure 3) and the peak values of these five VIs will therefore be used for RF modelling.



**Figure 3** Importance of the RF models variables for crop trait estimations and for yield prediction, expressed as drop-out loss of model performance (RMSE) for each variable related to the drop-out loss of the full model (dotted line). \*The RMSE values are in (%), (cm), (Mg DM ha<sup>-1</sup>), (Mg DM ha<sup>-1</sup>) and (Mg DM ha<sup>-1</sup>) respectively for the light interception, plant height, green leaf biomass, standing biomass and yield.

### 3.3. Machine learning model for crop traits estimation

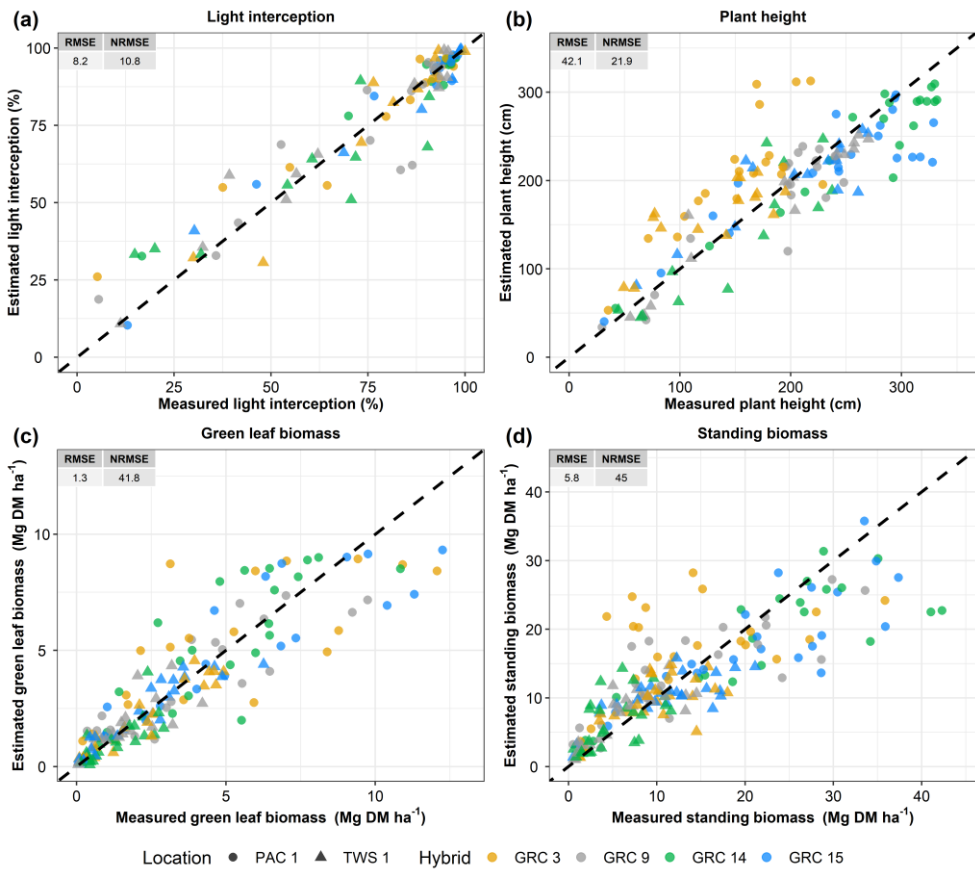
Crop traits distribution of contrasting *Miscanthus* hybrids measured at two locations is shown in Figure 4. The frequency distribution of the traits (light interception, plant height, green leaf biomass and standing biomass), which values were used for training and testing the models, showed that lower values were recorded in TWS 1 than in PAC 1 (Figure 4). For light interception, the range was from 5.2 % to 100 % in TWS 1 and PAC 1, and the mean was 58 % in TWS 1 and 81.5 % in PAC 1. The mean of plant height was 211 cm in PAC 1 and 147 cm in TWS 1, with the range from 28 cm to 344 cm and from 22 cm to 280 cm in PAC 1 and TWS 1, respectively. The range and the mean of green leaf biomass were from 0.14 Mg DM ha<sup>-1</sup> to 14.5 Mg DM ha<sup>-1</sup> and 5 Mg DM ha<sup>-1</sup> in PAC 1 and from 0.05 Mg DM ha<sup>-1</sup> to 6.3 Mg DM ha<sup>-1</sup> and 1.5 Mg DM ha<sup>-1</sup> in TWS 1. For standing biomass, the range was from 0.5 Mg DM ha<sup>-1</sup> to 46.4 Mg DM ha<sup>-1</sup> in PAC 1 and from 0.5 Mg DM ha<sup>-1</sup> to 21.1 Mg DM ha<sup>-1</sup> in TWS 1, and the mean was 8 Mg DM ha<sup>-1</sup> in TWS 1 and 18.9 Mg DM ha<sup>-1</sup> in PAC 1.



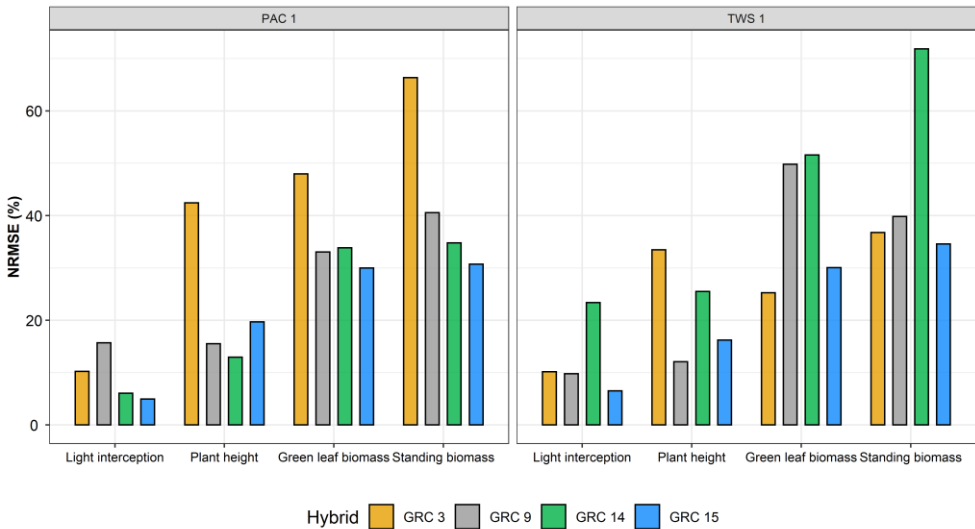
**Figure 4** Frequency distribution of *Miscanthus* traits at the two locations PAC 1 and TWS 1: (a) light interception (%), (b) plant height (cm), (c) green leaf biomass (Mg DM ha<sup>-1</sup>) and (d) standing biomass (Mg DM ha<sup>-1</sup>).

Overall, the RF model estimated crop traits with good model performances (Figure 5 and Figure 6). Among the crop traits, light interception was estimated with the greatest accuracy (NRMSE of 10.9 %, Figure 5). High model accuracy was also achieved for the estimation of plant height (21.8 NRMSE %), while the lowest model accuracies were observed for the green leaf biomass and standing biomass (42.2 % and 45.3 % of NRMSE, respectively Figure 5). For these last parameters, the RF model showed good accuracy from low to intermediate values, while above values of 5 Mg DM ha<sup>-1</sup> of green leaf biomass and 20 Mg DM ha<sup>-1</sup> of standing biomass the model performances dropped. The NRMSE performance metrics for each location and hybrid are reported in Figure 6. Generally, no relevant differences were observed between the two locations (Figure 6). In PAC 1, the

GRC 3 hybrid showed the worst performance for green leaf biomass, for standing biomass and for plant height, while in TWS 1, GRC 14 hybrid showed the worst performance for all the crop traits considered, except for plant height.



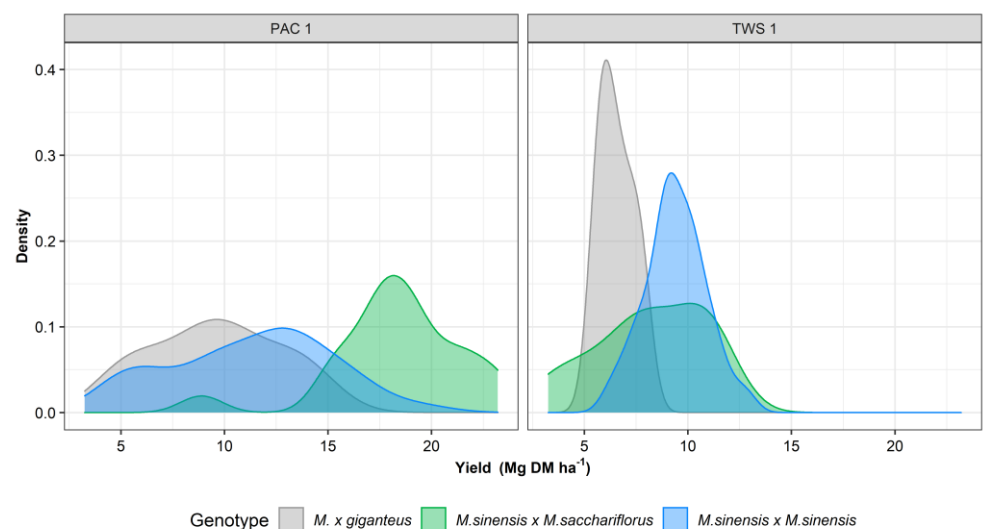
**Figure 5** Estimated vs measured crop traits on the ground of four *Miscanthus* hybrids growth at PAC 1 and TWS 1: (a) light interception (%), (b) plant height (cm), (c) green leaf biomass (Mg DM ha<sup>-1</sup>) and (d) standing biomass (Mg DM ha<sup>-1</sup>).



**Figure 6** NRMSE values of the RF models obtained for each crop trait assessed in four hybrids grown at the two locations PAC 1 and TWS 1. Note: lower values indicate higher estimation accuracies.

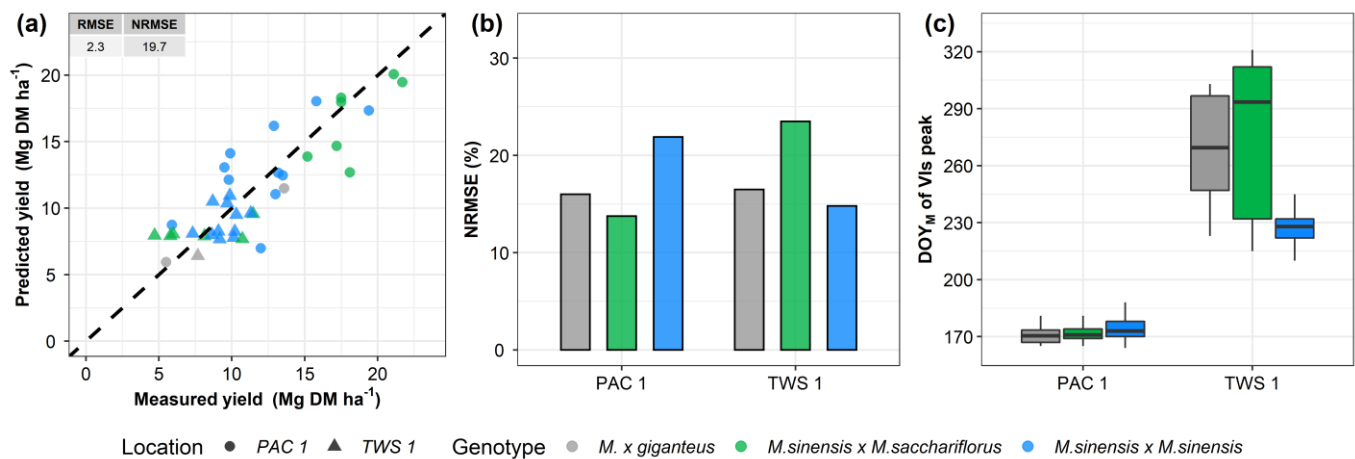
### 3.4. Machine learning model for yield prediction

The frequency distribution of yield measured for the 14 *Miscanthus* hybrids in PAC 1 and TWS 1 is shown in Figure 7. On average, the yield of *Miscanthus* was higher in PAC 1 than in TWS 1. In PAC 1, the highest yield was recorded by *M. sin x M. sac* (mean of 18.3 Mg DM ha<sup>-1</sup>). *M. sin x M. sin* productivity averaged 11.3 Mg DM ha<sup>-1</sup>, while *M. giganteus* was the less productive (mean of 9.6 Mg DM ha<sup>-1</sup>). In TWS 1, the highest yield was recorded by *M. sin x M. sin* (mean of 9.4 Mg DM ha<sup>-1</sup>). *M. sin x M. sac* productivity averaged 8.2 Mg DM ha<sup>-1</sup> while *M. giganteus* was the less productive (mean of 6.6 Mg DM ha<sup>-1</sup>). The RF model, trained and tested with the yield values reported in Figure 7, enabled a reliable prediction of *Miscanthus* yield for all hybrids using the peak derived from complete time series of VIs (Figure 8). The RF model obtained a RMSE value of 2.3 Mg DM ha<sup>-1</sup> and NRMSE value of 19.7 % (Figure 8a). In PAC 1, *M. sin x M. sac* showed the lowest NRMSE value while *M. sin x M. sin* showed the highest NRMSE value. On the contrary, in TWS 1, *M. sin x M. sin* showed the lowest NRMSE value while *M. sin x M. sac* showed the highest NRMSE value (Figure 8b). The modified days of the year (DOY<sub>M</sub>) of the peak of the VIs (greenWDRVI, NDVI, WDRVI, GNDVI and MTVI2) are reported in Figure 8c. On average, the VIs reached the peak earlier in PAC 1 (172 DOY<sub>M</sub> – June 20<sup>th</sup>) than in TWS 1 (DOY<sub>M</sub> 232– August 19<sup>th</sup>). In PAC 1, all hybrids reached the peak at the same time while in TWS 1 all *M. sin x M. sin* hybrids were the earliest to reach the peak (228 DOY<sub>M</sub> - August 15<sup>th</sup>), while *M. sin x M. sac* hybrids and *M. x giganteus* reached the peak along a wide timespan ranging from end-September until mid-November (Figure 8c).



**Figure 7** Frequency distribution of yield (Mg DM ha<sup>-1</sup>) for the different *Miscanthus* hybrids at the two locations PAC 1 and TWS 1.

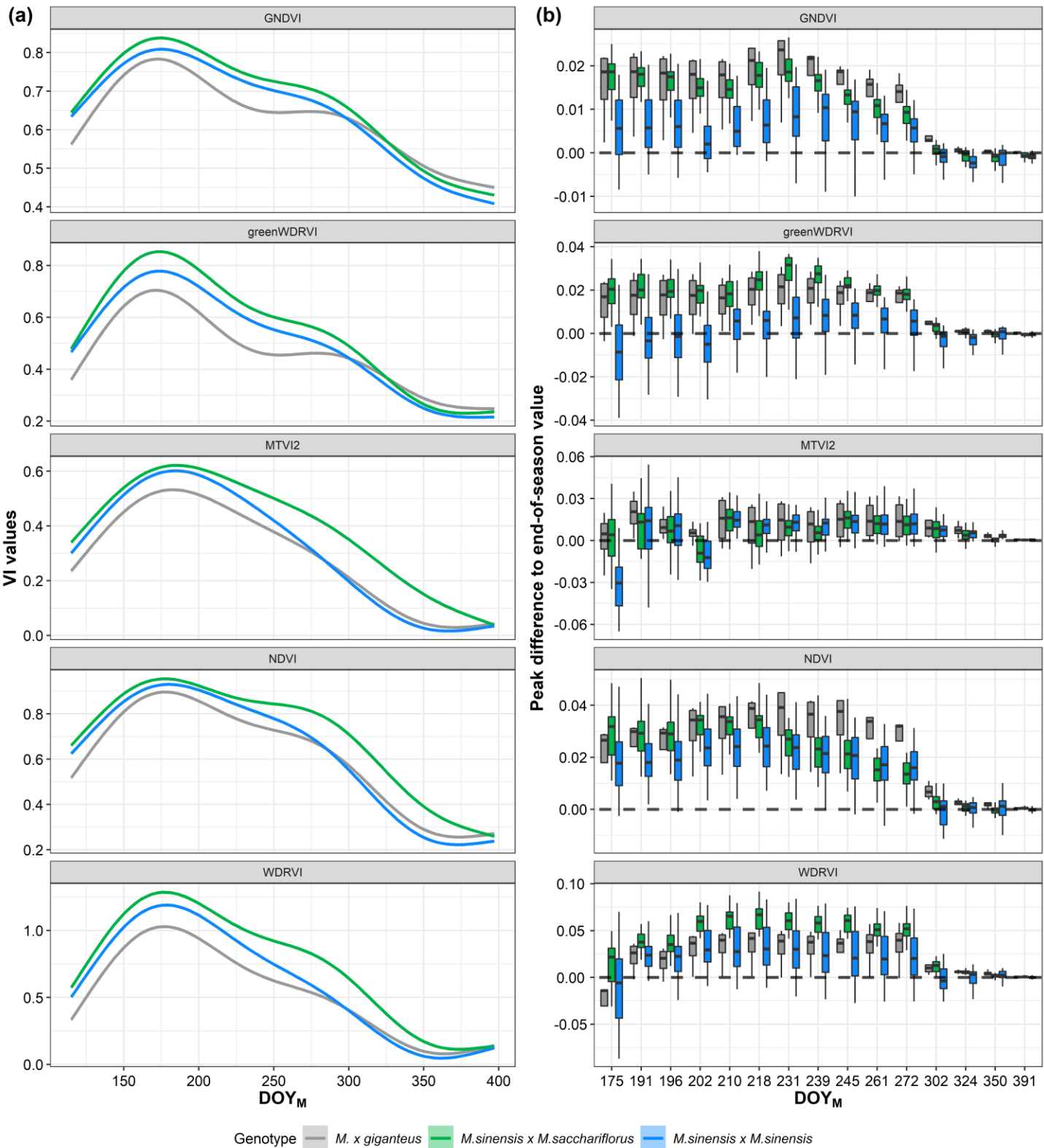




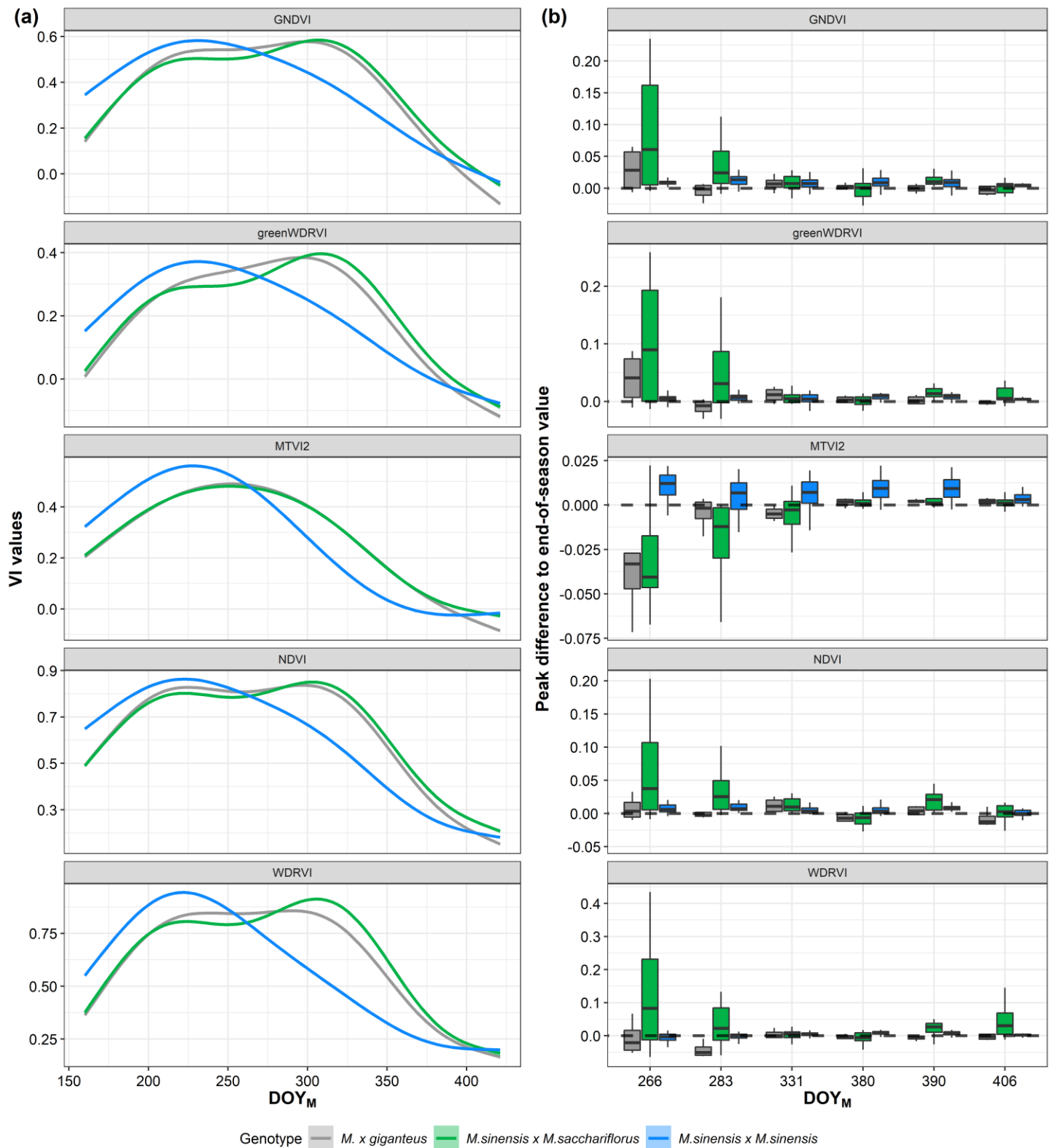
**Figure 8** (a) Predicted vs measured yield (Mg DM ha<sup>-1</sup>), (b) NRMSE of the RF model for yield prediction and (c) boxplot of the modified days of the year (DOY<sub>M</sub>) of the peak derived from complete time series of the 5 VIs (greenWDRVI, GNDVI, MTVI2, NDVI and WDRVI), of the different *Miscanthus* hybrids at two locations PAC 1 and TWS 1.

### 3.5. Time series of VIs and yield prediction analysis

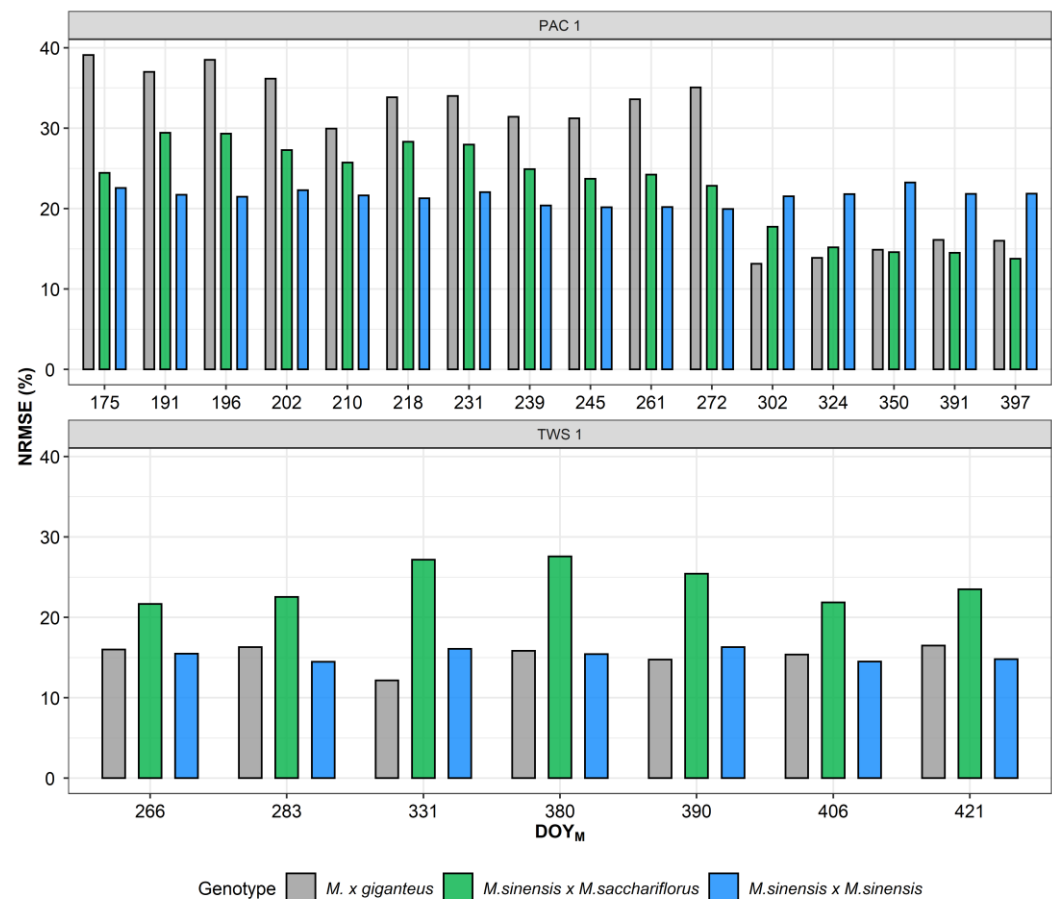
The complete time series of the five VIs identified as the most important for yield prediction (see § 3.2 and figure 3) are reported in Figure 9a for PAC 1 and Figure 10a for TWS 1. In PAC 1, all VIs values recorded throughout the growth of *Miscanthus* were the highest for *M. sin x M. sac* and the lowest for *M. giganteus* (Figure 9a), following the same order of the mean yield measured in the field (Figure 7). In TWS 1, similar time series of all VIs were recorded for the *M. sin x M. sac* and *M. giganteus* (Figure 10a). In particular, the peaks of *M. sin x M. sac* and of *M. giganteus* were reached later than that of *M. sin x M. sin*, as shown in Figure 8c. The variation throughout the season of the peak derived by fitting the VIs via generalized additive model (GAM) are displayed in Figure 9b for PAC 1 and Figure 10b for TWS 1. In PAC 1, the difference between the value of the peak derived from the complete time series of VIs and the value of the peak derived from the partial time series of VIs is the lowest (close to zero) after the DOY<sub>M</sub> 302 (end-October). Before this date, the peaks differences are lower for *M. sin x M. sin* than *M. sin x M. sac* and *M. giganteus* (Figure 9b). In TWS 1, the difference between the value of the peak derived from the complete time series of VIs and the value of the peak derived from the partial time series of VIs is lowest after DOY<sub>M</sub> 331 (end-November). As for PAC 1, the peaks differences are lower for *M. sin x M. sin* than *M. giganteus* and *M. sin x M. sac* (Figure 10b) in TWS 1 before 331 DOY<sub>M</sub>. The timeline of the performance of the RF model tested with the VIs peak from partial time series is reported in Figure 11. In PAC 1, the NRMSE decreased until DOY<sub>M</sub> 302 for *M. giganteus* and *M. sin x M. sac* while for *M. sin x M. sin*, it remained stable for all UAV flights performed from 175 DOY<sub>M</sub> and onward. In TWS 1, no relevant differences in NRMSE were observed between the UAV flights performed from 266 DOY<sub>M</sub> and onward.



**Figure 9** (a) Time series of the five VIs (GNDVI, greenWDRVI, MTVI2, NDVI, WDRVI) fitted via generalized additive model (GAM) throughout the growing season of *Miscanthus* in PAC 1. Modified days of the year (DOY<sub>M</sub>) were calculated by adding 365 to the DOY of the corresponding year from January on. (b) Variation of the peak of the VIs derived from complete time series of the VIs as compared to the peak of the VIs derived from partial time series of the VIs. In the x-axis are reported the DOY<sub>M</sub> of the UAV flights performed during the season in PAC 1. In the y-axis are reported the peak differences between the peak derived to the end of the season in PAC 1 (397 DOY<sub>M</sub>) and the peak derived from partial time series fitted until the DOY<sub>M</sub> of the UAV flight.



**Figure 10** (a) Time series of the five VIs (GNDVI, greenWDRVI, MTVI2, NDVI, WDRVI) fitted via generalized additive model (GAM) throughout the growing season of *Miscanthus* in TWS 1. Modified days of the year ( $DOY_M$ ) were calculated by adding 365 to the DOY of the corresponding year from January on. (b) Variation of the peak of the VIs derived from complete time series of the VIs as compared to the peak of the VIs derived from partial time series of the VIs. In the x-axis are reported the  $DOY_M$  of the UAV flights performed during the season in TWS 1. In the y-axis are reported the peak differences between the peak derived to the end of the season in TWS 1 (421  $DOY_M$ ) and the peak derived from partial time series fitted until the  $DOY_M$  of the UAV flight.



**Figure 11** NRMSE values of the RF model for yield prediction, trained with the peak of the VIs derived from complete time series and tested with the peak of the VIs derived from partial time series fitted until the modified days of the year (DOY<sub>M</sub>) of the UAV flight, at the two locations PAC 1 and TWS 1.

#### 4. Discussion

The use of UAV-based remote sensing provides a great potential for high-throughput phenotyping (HTP) at plot scale with applications in both breeding and in estimating the quality and quantity of the biomass for optimising downstream management of biomass fluxes. In this study, vegetation indices (VIs) and their peak were derived from unmanned aerial vehicles (UAVs) mounted multispectral sensors, to estimate crop traits (light interception, plant height, green leaf biomass and standing biomass) and to predict the final harvestable yield of novel *Miscanthus* hybrids and common *M. x giganteus* grown at two sites (Italy and UK).

##### 4.1. The importance of VIs intercalibration procedure for multi-sensor interoperability

Intercalibrating VIs of multi-sensor is relevant for remote sensing crop monitoring [54], particularly when the objective is to build models to estimate crop traits or to predict yield, and when sensors with different spectral characteristics are used. Indeed, the models might not reach the same accuracy if the VIs are calculated with sensors with different spectral band characteristics. The spectral signatures simulated from the PROSAIL model were used in this study to intercalibrate VIs calculated from two common multispectral cameras (MicaSense RedEdge-MX and SlantRange 4P). Ideally, the VIs from any sensor can be then intercalibrated against the VIs of a selected reference sensor. This intercalibration approach is commonly adopted for VIs obtained from different satellites [59] but to our knowledge, this is the first time that such an approach is applied to UAV multispectral sensors. PROSAIL model was used to simulate the canopy reflectance based on the value of specific *Miscanthus* traits, such as LAI, chlorophyll content, dry matter content, relative

equivalent water thickness, leaf inclination distribution and two site-specific info: solar zenith angle and relative azimuth angle. This approach was applied to the values of 15 VIs. EVI2, MSAVI and SAVI were the three VIs with the closest 1:1 relation, which indicates that these VIs have a very low sensitivity to the spectral characteristics of the two sensors used in the present work. Similar results with no need for intercalibration for the same VIs were reported by Li *et al.* [87] that compared ETM+ and OLI satellite imageries. However, the other VIs evaluated in this study (Datt1, GNDVI, GOSAVI, greenWDRVI, MTVI1, MTVI2, NDRE, NDVI, OSAVI, OSAVI2, rededgeWDRVI and WDRVI) showed a higher sensitivity to sensor characteristics, underlining the importance of intercalibrating VIs for multi-sensor interoperability. In particular, the VIs based on the green and red-edge bands showed a higher variation than the VIs based on the red band at all ranges of values. These differences are explained by the differences in the spectral characteristics of MicaSense and SlantRange in the green and red-edge bands. In particular, the SlantRange sensor has a broader green FWHM (100 nm) and a different central wavelength (710 nm) of the red edge compared to the MicaSense sensor (green FWHM: 27 nm and red edge central wavelength: 717 nm). This difference was already reported to cause considerable signal differences in other studies [88,89].

This study highlighted the importance of intercalibrating different multispectral sensors to increase interoperability in remote sensing. Kim *et al.* [90] and Villaescusa-Nadal *et al.* [91] reported that the use of linear regression equations to intercalibrate multi-sensor contributes to significantly correct (up to 50%) the effects of different spectral characteristics on VIs. However, it would be interesting to validate this procedure by flying simultaneously on the same field with two UAV sensors, comparing the values of the VIs with and without the intercalibration procedure, in order to evaluate the improvement in terms of VIs compatibility and multi-sensor interoperability. In fact, even if the spectral characteristics of the multispectral sensor are the factor that influences the most the compatibility of the VIs of different sensors [49], other factors cause differences of VIs such as the atmospheric conditions during acquisition [48]. In the UAV images acquisitions, the changing light and meteorological conditions during the flights can affect the quality of the spectral data [92]. Therefore, this procedure is limited by other factors that cannot be considered by applying a simulated regression coefficient. However, using equations able to intercalibrate VIs derived by multiple multispectral sensors can reduce differences in VIs improving crop monitoring and modelling for estimation of crop traits and prediction of yield.

#### 4.2. Estimating *Miscanthus* traits with machine learning

This study estimated *Miscanthus* traits using the random forest (RF) machine learning algorithm. The RF model was trained with the data collected on three novel seed-based *Miscanthus* hybrids and the common rhizome-based genotype *M. x giganteus*, at two contrasting locations (North-West Italy and Mid-West Wales). The RF algorithm, using 15 common VIs, successfully estimated crop traits, solving the non-linear responses between VIs and crop traits observed by Li *et al.* [31] for other perennial crops. The estimation of the crop traits from time series of VIs acquired by UAV-based remote sensing can generate more data useful to calibrate existing *Miscanthus* crop models and to re-fine these models for novel *Miscanthus* hybrids in contrasting environments. The crop trait estimated with the greatest accuracy was light interception that showed a RMSE of 8.4 %, the accuracy being especially good at high values of light interception (Figure 5). This result is in agreement with Guillen-Climent *et al.* [93] who found that the fraction of intercepted photosynthetically active radiation (fIPAR) was successfully estimated by a ML algorithm. Upreti *et al.* [33] found similar values of NRMSE (12.06 %) using the RF tree bagger approach for estimating the fraction of absorbed photosynthetically active radiation (fAPAR) of durum wheat. Good model accuracy was also achieved for plant height estimation (RMSE = 42 cm and NRMSE = 21.8 %, Figure 5). A similar RMSE value (41 cm) was found by Han *et al.* [34] for plant height estimation of maize using the crop surface model, and by Tao *et al.* [94] (NRMSE = 21.2 %) in the estimation of plant height of winter wheat using UAV hyperspectral images. The worst model accuracy was found for the green leaf biomass and



standing biomass with 1.3 Mg DM ha<sup>-1</sup> and 5.8 Mg DM ha<sup>-1</sup> of RMSE, respectively (Figure 5). However, for these parameters, the model showed good accuracy from low to intermediate values, while above values of 5 Mg DM ha<sup>-1</sup> of green leaf biomass and 20 Mg DM ha<sup>-1</sup> of standing biomass, the model performance dropped. The model could be affected by errors in the estimation in these intervals due to fewer data used to train the model [95], in fact, most of them were collected in PAC 1 (Figure 4). The different levels of accuracy of the models in estimating light interception and plant height compared to green leaf biomass and standing biomass could also be related to the period in which the measurements were taken and to the response of VIs during senescence. Field measurements of light interception and plant height were carried out in each environment from emergence until they peaked, which explains their good model accuracy. Estimation of crop traits with data coming from single UAV flights across the growing seasons is particularly affected after biomass reaches its maximum value in autumn. After biomass peak, with the start of the senescence period, the values of VIs start to decrease (Figure 9a and Figure 10a) [96] while *Miscanthus* green leaf biomass and standing biomass values remained stable or slightly decreased during this period. The difference of rate of decrease between VIs and crop traits during senescence is a key aspect to consider in remote sensing estimation of crop traits. The importance of the senescence stage in the crop traits estimation is confirmed by the results on the variables' importance (Figure 3). Indeed, the phenological variable "Stage" was the most important variable in the estimation of plant height, green leaf biomass and standing biomass. The plant height of the GRC 3 (a *M. sin x M. sin* planted at high density) were poorly estimated, and this could be due to its canopy architecture and flowering time. *M. sin x M. sin* hybrid has many distinguishable stems flowering (where plant height is measured) but leaves are particularly curved and attached along the stem at a lower height than *M. sin x M. sac*. In addition, this genotype was transplanted at higher densities (3 plants m<sup>-2</sup>) and flowered earlier (end of August) compared to other genotypes that flowered in early autumn. This more "prostrate" canopy architecture (with a higher stem segment between inflorescences and bent leaves) introduced noise in the plant height estimation from UAVs [97] since most of the reflectance comes from bent leaf mass. This noise caused by changes in plant architecture and the onset of flower can be seen in the NRMSE values of the RF models at the PAC 1 site (Figure 6). In fact, the earlier a genotype with prostrate architecture flowers, the worse is the estimation of plant height and biomass from UAV.

#### 4.3. Yield prediction using machine learning and peak of VIs

The random forest (RF) trained with the peak derived from complete time series of five VIs acquired by UAVs was able to predict the yield of the 14 *Miscanthus* hybrids. The peak value of greenWDRVI, NDVI, WDRVI, GNDVI and MTVI2 resulted the most important features for RF modelling as derived from previous analysis (see § 3.2, Figure 3). The RF model accurately predicted the yield with 2.3 Mg DM ha<sup>-1</sup> of RMSE and 19.7 % NRMSE (Figure 8a). The peak for *Miscanthus* hybrids occurs on average in mid-summer and early autumn in southern/warm (Italy) and northern/cold (UK) locations, respectively. The importance of the peak as land surface phenology (LSP) descriptors for yield prediction was already reported by Prasad *et al.* [98], who found that peak had the highest correlation with cotton yield prediction compared to other LSP descriptors. Similar results were reported by Montazeaud *et al.* [47] who found a high correlation between the peak of NDVI and the yield, and by Liu *et al.* [20], who found that the EVI2 peak was a good predictor of grain yield. Among the VIs used in the RF model, the peak of the VIs based on the green band as GNDVI and greenWDRVI (Figure 3) were the most important variables for predicting not only *Miscanthus* yield at harvest but also to estimate standing biomass and green leaf biomass during the growing season (Figure 4). Similar results for GNDVI were found in switchgrass and other warm-season perennial grasses [29]. In order to assess the capability of the model to predict the yield months before harvest (i.e. using only early season UAV acquisition and not waiting to perform UAV monitoring along the entire crop season), RF model was calculated using the peak derived from partial time series of VIs and performance analysed. In the UK, the RF model accurately predicted the

yield five months before harvest for all *Miscanthus* hybrids. In Italy, the yield of the *M. sin x M. sin* hybrids can be predicted with good accuracy seven months before harvest while *M. giganteus* and *M. sin x M. sac* hybrids required more time, as a good accuracy was obtained 3-4 months before harvest. This capability of the model to predict yield months before harvest is possible if no damage to plants later in the season occurs, thus considering a steady development process. Given the good accuracy achieved by the RF model even when peak value is assessed on a limited number of data (see § 3.5, Figure 11), this approach can be considered as a suitable method to predict yield several months before harvest. However, it is possible that no time series analyses (i.e. a complete series of dense observation from sowing to harvest) are required, and the 5 VIs values could be used directly as inputs for yield prediction modelling if the UAV data is acquired in the range of expected peak occurrence. This opportunity was also reported for switchgrass [29] using multispectral images acquired from multiple UAV flights, underlying the possibility of reducing the number of multispectral observations (e.g. 25 UAV flights were performed in PAC 1) to 2-3 UAV flights in proximity of the VIs peak period. In this context, the identification of VIs peak period, which is generally related to crop physiology and phenology (reach the maximum production), is important in term of cost (reduction of number of observations) and of operability of the yield prediction model. This model operability, intended as a capability of the model to accurately predict the yield some months before the harvest, is extremely relevant for improving the logistics of biomass supply chain of *Miscanthus* and for supporting the improvement of crop modelling with remote sensing data.

## 5. Conclusions

This study demonstrated that vegetation indices (VIs) derived from unmanned aerial vehicle (UAV) multispectral images acquired in Italy and UK can be successfully used in random forest (RF) machine learning (ML) algorithm to estimate the light interception, plant height, green leaf biomass and standing biomass, and to predict the yield of novel *Miscanthus* hybrids using the peak derived from VIs time series. This study evaluated the timeline of the performance of the model using peak derived from partial VIs time series and the RF model showed a good capability to predict the yield months before the harvest both in Italy and in the UK by using a limited number of UAV observation. In particular, the results suggested that the VIs values acquired during the peak period (without using complete or partial time series) could be used directly for yield prediction, increasing the model operability. Yield prediction and high-throughput phenotyping based on ML algorithms and on UAV remote sensing can improve the logistics of biomass supply chain, for supporting breeding programs, and for improving crop modelling of novel *Miscanthus* hybrids. UAV platforms are suitable tools for HTP applications, as they enable the monitoring of small plots or field scale trials with numerous genotypes, due to their ability to capture high-resolution images. However, the satellite platforms are more suited for yield prediction, as they can collect data of many fields simultaneously and can develop applications to predict commercial yield at regional and national scales. In addition, this study reported for the first time a methodology to overcome the issue of multi-sensor interoperability among UAV multispectral sensors. The use of intercalibrating equations derived from the PROSAIL model proved to be a powerful tool to intercalibrate VIs from multi-sensor with different spectral characteristics. Although this intercalibration procedure is relevant for the upscale of models from experimental plots to field by intercalibrating the UAV with satellites sensor characteristics, it is limited because it only considers the spectral sensor characteristics and no other factors such as light and meteorological conditions during the flights, which may affect the quality of the spectral data, and which cannot be considered by applying a simulated regression coefficient.

**Author Contributions:** Conceptualization, G.I., M.C., A.F. and S.A.; methodology, G.I. and M.C.; software, G.I. and M.C.; formal analysis, G.I.; data curation, G.I.; visualization, G.I.; investigation, G.I., M.C., J.B., A.M., H.B., E.M., C.A. and D.A.; resources, J.K., L.M.T., J.C.B., A.K. and S.A.;

writing—original draft preparation, G.I.; writing—review and editing, A.F., J.C.B., M.B. and S.A.; funding acquisition, S.A.. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study is part of the GRACE project which has received funding from the Bio-based Industries Joint Undertaking (JU) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 745012. The JU receives support from the European Union's Horizon 2020 research and innovation programme and the Bio-based Industries Consortium.

**Data Availability Statement:** The data that support the findings of this work are available from the corresponding author upon reasonable request.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Lewandowski, I.; Clifton-Brown, J.C.; Scurlock, J.M.O.; Huisman, W. Miscanthus: European Experience with a Novel Energy Crop. *Biomass and Bioenergy* **2000**, *19*, 209–227, doi:10.1016/S0961-9534(00)00032-5.
2. Clifton-Brown, J.; Hastings, A.; Mos, M.; McCalmont, J.P.; Ashman, C.; Awty-Carroll, D.; Cerazy, J.; Chiang, Y.-C.; Cosentino, S.; Cracroft-Eley, W.; et al. Progress in Upscaling *Miscanthus* Biomass Production for the European Bio-Economy with Seed-Based Hybrids. *GCB Bioenergy* **2017**, *9*, 6–17, doi:10.1111/gcbb.12357.
3. Clifton-Brown, J.C.; Lewandowski, I.; Andersson, B.; Basch, G.; Christian, D.G.; Kjeldsen, J.B.; Jørgensen, U.; Mortensen, J. V.; Riche, A.B.; Schwarz, K.-U.; et al. Performance of 15 *Miscanthus* Genotypes at Five Sites in Europe. *Agron. J.* **2001**, *93*, 1013–1019, doi:10.2134/agronj2001.9351013x.
4. Jones, M.B.; Zimmermann, J.; Clifton-Brown, J. Long-Term Yields and Soil Carbon Sequestration from *Miscanthus*: A Review. In *Perennial Biomass Crops for a Resource-Constrained World*; Springer International Publishing: Cham, 2016; pp. 43–49.
5. CLIFTON-BROWN, J.C.; BREUER, J.; JONES, M.B. Carbon Mitigation by the Energy Crop, *Miscanthus*. *Glob. Chang. Biol.* **2007**, *13*, 2296–2307, doi:10.1111/j.1365-2486.2007.01438.x.
6. Jeżowski, S. Yield Traits of Six Clones of *Miscanthus* in the First 3 Years Following Planting in Poland. *Ind. Crops Prod.* **2008**, *27*, 65–68, doi:10.1016/j.indcrop.2007.07.013.
7. Larsen, S.U.; Jørgensen, U.; Kjeldsen, J.B.; Lærke, P.E. Long-Term *Miscanthus* Yields Influenced by Location, Genotype, Row Distance, Fertilization and Harvest Season. *BioEnergy Res.* **2014**, *7*, 620–635, doi:10.1007/s12155-013-9389-1.
8. Clifton-Brown, J.; Harfouche, A.; Casler, M.D.; Dylan Jones, H.; Macalpine, W.J.; Murphy-Bokern, D.; Smart, L.B.; Adler, A.; Ashman, C.; Awty-Carroll, D.; et al. Breeding Progress and Preparedness for Mass-scale Deployment of Perennial Lignocellulosic Biomass Crops Switchgrass, *Miscanthus*, Willow and Poplar. *GCB Bioenergy* **2019**, *11*, 118–151, doi:10.1111/gcbb.12566.
9. Lewandowski, I.; Clifton-Brown, J.; Trindade, L.M.; van der Linden, G.C.; Schwarz, K.-U.U.; Müller-Sämann, K.; Anisimov, A.; Chen, C.-L.L.; Dolstra, O.; Donnison, I.S.; et al. Progress on Optimizing *Miscanthus* Biomass Production for the European Bioeconomy: Results of the EU FP7 Project OPTIMISC. *Front. Plant Sci.* **2016**, *7*, 1620, doi:10.3389/fpls.2016.01620.
10. van der Cruysen, K.; Al Hassan, M.; van Erven, G.; Dolstra, O.; Trindade, L.M. Breeding Targets to Improve Biomass Quality in *Miscanthus*. *Molecules* **2021**, *26*, 254, doi:10.3390/molecules26020254.
11. Clifton-Brown, J.; Schwarz, K.-U.; Awty-Carroll, D.; Iurato, A.; Meyer, H.; Greef, J.; Gwyn, J.; Mos, M.; Ashman, C.; Hayes, C.; et al. Breeding Strategies to Improve *Miscanthus* as a Sustainable Source of Biomass for Bioenergy and Biorenewable Products. *Agronomy* **2019**, *9*, 673, doi:10.3390/agronomy9110673.
12. Hastings, A.; Mos, M.; Yesufu, J.A.; McCalmont, J.; Schwarz, K.; Shafei, R.; Ashman, C.; Nunn, C.; Schuele, H.; Cosentino, S.; et al. Economic and Environmental Assessment of Seed and Rhizome Propagated *Miscanthus* in the UK. *Front. Plant Sci.* **2017**, *8*, 1058, doi:10.3389/fpls.2017.01058.
13. Pancaldi, F.; Trindade, L.M. Marginal Lands to Grow Novel Bio-Based Crops: A Plant Breeding Perspective. *Front. Plant Sci.* **2020**, *11*, doi:10.3389/fpls.2020.00227.

14. de Wit, A.J.W.; van Diepen, C.A. Crop Growth Modelling and Crop Yield Forecasting Using Satellite-Derived Meteorological Inputs. *Int. J. Appl. Earth Obs. Geoinf.* **2008**, *10*, 414–425, doi:10.1016/j.jag.2007.10.004.
15. Keating, B.; Carberry, P.; Hammer, G.; Probert, M.; Robertson, M.; Holzworth, D.; Huth, N.; Hargreaves, J.N.; Meinke, H.; Hochman, Z.; et al. An Overview of APSIM, a Model Designed for Farming Systems Simulation. *Eur. J. Agron.* **2003**, *18*, 267–288, doi:10.1016/S1161-0301(02)00108-9.
16. *Decision Support Systems in Potato Production*; MacKerron, D.K.L., Haverkort, A.J., Eds.; Wageningen Academic Publishers: The Netherlands, 2004; ISBN 978-90-76998-30-5.
17. HASTINGS, A.; CLIFTON-BROWN, J.; WATTENBACH, M.; MITCHELL, C.P.; SMITH, P. The Development of MISCANFOR, a New Miscanthus Crop Growth Model: Towards More Robust Yield Predictions under Different Climatic and Soil Conditions. *GCB Bioenergy* **2009**, *1*, 154–170, doi:10.1111/j.1757-1707.2009.01007.x.
18. Zhang, B.; Hastings, A.; Clifton-Brown, J.C.; Jiang, D.; Faaij, A.P.C. Modeled Spatial Assessment of Biomass Productivity and Technical Potential of Miscanthus × Giganteus, Panicum Virgatum L., and Jatropha on Marginal Land in China. *GCB Bioenergy* **2020**, *12*, 328–345, doi:10.1111/gcbb.12673.
19. Henner, D.N.; Hastings, A.; Pogson, M.; McNamara, N.P.; Davies, C.A.; Smith, P. PopFor: A New Model for Estimating Poplar Yields. *Biomass and Bioenergy* **2020**, *134*, 105470, doi:10.1016/j.biombioe.2020.105470.
20. Liu, J.; Shang, J.; Qian, B.; Huffman, T.; Zhang, Y.; Dong, T.; Jing, Q.; Martin, T. Crop Yield Estimation Using Time-Series MODIS Data and the Effects of Cropland Masks in Ontario, Canada. *Remote Sens.* **2019**, *11*, 2419, doi:10.3390/rs11202419.
21. Richter, G.M.; Agostini, F.; Barker, A.; Costomiris, D.; Qi, A. Assessing On-Farm Productivity of Miscanthus Crops by Combining Soil Mapping, Yield Modelling and Remote Sensing. *Biomass and Bioenergy* **2016**, *85*, 252–261, doi:10.1016/j.biombioe.2015.12.024.
22. Impollonia, G.; Croci, M.; Martani, E.; Ferrarini, A.; Kam, J.; Trindade, L.M.; Clifton-Brown, J.; Amaducci, S. Moisture Content Estimation and Senescence Phenotyping of Novel Miscanthus Hybrids Combining UAV-based Remote Sensing and Machine Learning. *GCB Bioenergy* **2022**, *14*, 639–656, doi:10.1111/gcbb.12930.
23. Potgieter, A.B.; George-Jaeggli, B.; Chapman, S.C.; Laws, K.; Suárez Cadavid, L.A.; Wixted, J.; Watson, J.; Eldridge, M.; Jordan, D.R.; Hammer, G.L. Multi-Spectral Imaging from an Unmanned Aerial Vehicle Enables the Assessment of Seasonal Leaf Area Dynamics of Sorghum Breeding Lines. *Front. Plant Sci.* **2017**, *8*, 1532, doi:10.3389/fpls.2017.01532.
24. Blancon, J.; Dutartre, D.; Tixier, M.-H.H.; Weiss, M.; Comar, A.; Praud, S.; Baret, F. A High-Throughput Model-Assisted Method for Phenotyping Maize Green Leaf Area Index Dynamics Using Unmanned Aerial Vehicle Imagery. *Front. Plant Sci.* **2019**, *10*, 685, doi:10.3389/fpls.2019.00685.
25. Antonucci, G.; Impollonia, G.; Croci, M.; Potenza, E.; Marcone, A.; Amaducci, S. Evaluating Biostimulants via High-Throughput Field Phenotyping: Biophysical Traits Retrieval through PROSAIL Inversion. *Smart Agric. Technol.* **2023**, *3*, 100067, doi:10.1016/j.atech.2022.100067.
26. Jongschaap, R.E.E. Run-Time Calibration of Simulation Models by Integrating Remote Sensing Estimates of Leaf Area Index and Canopy Nitrogen. *Eur. J. Agron.* **2006**, *24*, 316–324, doi:10.1016/j.eja.2005.10.009.
27. Prévot, L.; Chauki, H.; Troufleau, D.; Weiss, M.; Baret, F.; Brisson, N. Assimilating Optical and Radar Data into the STICS Crop Model for Wheat. *Agronomie* **2003**, *23*, 297–303, doi:10.1051/agro:2003003.
28. Ferchichi, A.; Abbes, A. Ben; Barra, V.; Farah, I.R. Forecasting Vegetation Indices from Spatio-Temporal Remotely Sensed Data Using Deep Learning-Based Approaches: A Systematic Literature Review. *Ecol. Inform.* **2022**, 101552, doi:10.1016/j.ecoinf.2022.101552.
29. Hamada, Y.; Zumpf, C.R.; Cacho, J.F.; Lee, D.; Lin, C.-H.; Boe, A.; Heaton, E.; Mitchell, R.; Negri, M.C.; Rescia, A.; et al. Remote Sensing-Based Estimation of Advanced Perennial Grass Biomass Yields for Bioenergy. *Land* **2021**, *10*, 1221, doi:10.3390/land10111221.
30. Wang, J.; Badenhorst, P.; Phelan, A.; Pembleton, L.; Shi, F.; Cogan, N.; Spangenberg, G.; Smith, K. Using Sensors and Unmanned Aircraft Systems for High-Throughput Phenotyping of Biomass in Perennial Ryegrass Breeding Trials. *Front.*



*Plant Sci.* **2019**, *10*, doi:10.3389/fpls.2019.01381.

31. Li, F.; Piasecki, C.; Millwood, R.J.; Wolfe, B.; Mazarei, M.; Stewart, C.N. High-Throughput Switchgrass Phenotyping and Biomass Modeling by UAV. *Front. Plant Sci.* **2020**, *11*, 1532, doi:10.3389/fpls.2020.574073.
32. Jin, S.; Su, Y.; Gao, S.; Hu, T.; Liu, J.; Guo, Q. The Transferability of Random Forest in Canopy Height Estimation from Multi-Source Remote Sensing Data. *Remote Sens.* **2018**, *10*, 1183, doi:10.3390/rs10081183.
33. Upreti, D.; Huang, W.; Kong, W.; Pascucci, S.; Pignatti, S.; Zhou, X.; Ye, H.; Casa, R. A Comparison of Hybrid Machine Learning Algorithms for the Retrieval of Wheat Biophysical Variables from Sentinel-2. *Remote Sens.* **2019**, *11*, 481, doi:10.3390/rs11050481.
34. Han, L.; Yang, G.; Dai, H.; Xu, B.; Yang, H.; Feng, H.; Li, Z.; Yang, X. Modeling Maize Above-Ground Biomass Based on Machine Learning Approaches Using UAV Remote-Sensing Data. *Plant Methods* **2019**, *15*, 10, doi:10.1186/s13007-019-0394-z.
35. Adam, E.; Mutanga, O.; Abdel-Rahman, E.M.; Ismail, R. Estimating Standing Biomass in Papyrus ( *Cyperus Papyrus* L.) Swamp: Exploratory of in Situ Hyperspectral Indices and Random Forest Regression. *Int. J. Remote Sens.* **2014**, *35*, 693–714, doi:10.1080/01431161.2013.870676.
36. Verrelst, J.; Malenovsky, Z.; Van der Tol, C.; Camps-Valls, G.; Gastellu-Etchegorry, J.-P.; Lewis, P.; North, P.; Moreno, J. Quantifying Vegetation Biophysical Variables from Imaging Spectroscopy Data: A Review on Retrieval Methods. *Surv. Geophys.* **2019**, *40*, 589–629, doi:10.1007/s10712-018-9478-y.
37. Wang, L.; Zhou, X.; Zhu, X.; Dong, Z.; Guo, W. Estimation of Biomass in Wheat Using Random Forest Regression Algorithm and Remote Sensing Data. *Crop J.* **2016**, *4*, 212–219, doi:10.1016/j.cj.2016.01.008.
38. Belgiu, M.; Drăguț, L. Random Forest in Remote Sensing: A Review of Applications and Future Directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31, doi:10.1016/j.isprsjprs.2016.01.011.
39. Hunt, M.L.; Blackburn, G.A.; Carrasco, L.; Redhead, J.W.; Rowland, C.S. High Resolution Wheat Yield Mapping Using Sentinel-2. *Remote Sens. Environ.* **2019**, *233*, 111410, doi:10.1016/j.rse.2019.111410.
40. Jeong, J.H.; Resop, J.P.; Mueller, N.D.; Fleisher, D.H.; Yun, K.; Butler, E.E.; Timlin, D.J.; Shim, K.-M.; Gerber, J.S.; Reddy, V.R.; et al. Random Forests for Global and Regional Crop Yield Predictions. *PLoS One* **2016**, *11*, e0156571, doi:10.1371/journal.pone.0156571.
41. Marques Ramos, A.P.; Prado Osco, L.; Elis Garcia Furuya, D.; Nunes Gonçalves, W.; Cordeiro Santana, D.; Pereira Ribeiro Teodoro, L.; Antonio da Silva Junior, C.; Fernando Capristo-Silva, G.; Li, J.; Henrique Rojo Baio, F.; et al. A Random Forest Ranking Approach to Predict Yield in Maize with Uav-Based Vegetation Spectral Indices. *Comput. Electron. Agric.* **2020**, *178*, 105791, doi:10.1016/j.compag.2020.105791.
42. Senthilnath, J.; Dokania, A.; Kandukuri, M.; K.N., R.; Anand, G.; Omkar, S.N.N. Detection of Tomatoes Using Spectral-Spatial Methods in Remotely Sensed RGB Images Captured by UAV. *Biosyst. Eng.* **2016**, *146*, 16–32, doi:10.1016/j.biosystemseng.2015.12.003.
43. de Beurs, K.M.; Henebry, G.M. Land Surface Phenology and Temperature Variation in the International Geosphere-Biosphere Program High-Latitude Transects. *Glob. Chang. Biol.* **2005**, *11*, 779–790, doi:10.1111/j.1365-2486.2005.00949.x.
44. Ji, Z.; Pan, Y.; Zhu, X.; Wang, J.; Li, Q. Prediction of Crop Yield Using Phenological Information Extracted from Remote Sensing Vegetation Index. *Sensors* **2021**, *21*, 1406, doi:10.3390/s21041406.
45. Meroni, M.; d'Andrimont, R.; Vrieling, A.; Fasbender, D.; Lemoine, G.; Rembold, F.; Seguini, L.; Verhegghen, A. Comparing Land Surface Phenology of Major European Crops as Derived from SAR and Multispectral Data of Sentinel-1 and -2. *Remote Sens. Environ.* **2021**, *253*, 112232, doi:10.1016/J.RSE.2020.112232.
46. de Beurs, K.M.; Henebry, G.M. Spatio-Temporal Statistical Methods for Modelling Land Surface Phenology. In *Phenological Research*; Springer Netherlands: Dordrecht, 2010; pp. 177–208.
47. Montazeaud, G.; Karatoğma, H.; Öztürk, I.; Roumet, P.; Ecartot, M.; Crossa, J.; Özer, E.; Özdemir, F.; Lopes, M.S. Predicting Wheat Maturity and Stay-Green Parameters by Modeling Spectral Reflectance Measurements and Their Contribution to Grain Yield under Rainfed Conditions. *F. Crop. Res.* **2016**, *196*, 191–198, doi:10.1016/j.fcr.2016.06.021.



48. Psomiadis, E.; Dercas, N.; Dalezios, N.R.; Spiropoulos, N. V. Evaluation and Cross-Comparison of Vegetation Indices for Crop Monitoring from Sentinel-2 and Worldview-2 Images. In Proceedings of the Remote Sensing for Agriculture, Ecosystems, and Hydrology XIX; Neale, C.M., Maltese, A., Eds.; SPIE, November 2 2017; p. 79.
49. Théau, J.; Sankey, T.T.; Weber, K.T. Multi-Sensor Analyses of Vegetation Indices in a Semi-Arid Environment. *GIScience Remote Sens.* **2010**, *47*, 260–275, doi:10.2747/1548-1603.47.2.260.
50. Hoque, M.A.-A.; Phinn, S. Methods for Linking Drone and Field Hyperspectral Data to Satellite Data. In *Fundamentals, Sensor Systems, Spectral Libraries, and Data Mining for Vegetation*; CRC Press, 2018; pp. 321–354.
51. Emilien, A.-V.; Thomas, C.; Thomas, H. UAV & Satellite Synergies for Optical Remote Sensing Applications: A Literature Review. *Sci. Remote Sens.* **2021**, *3*, 100019, doi:10.1016/j.srs.2021.100019.
52. Brown, M.E.; Pinzon, J.E.; Didan, K.; Morisette, J.T.; Tucker, C.J. Evaluation of the Consistency of Long-Term NDVI Time Series Derived from AVHRR, SPOT-Vegetation, SeaWiFS, MODIS, and Landsat ETM+ Sensors. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 1787–1793, doi:10.1109/TGRS.2005.860205.
53. GALLO, K.; JI, L.; REED, B.; EIDENSHINK, J.; DWYER, J. Multi-Platform Comparisons of MODIS and AVHRR Normalized Difference Vegetation Index Data. *Remote Sens. Environ.* **2005**, *99*, 221–231, doi:10.1016/j.rse.2005.08.014.
54. Meroni, M.; Atzberger, C.; Vancutsem, C.; Gobron, N.; Baret, F.; Lacaze, R.; Eerens, H.; Leo, O. Evaluation of Agreement Between Space Remote Sensing SPOT-VEGETATION FAPAR Time Series. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 1951–1962, doi:10.1109/TGRS.2012.2212447.
55. She, X.; Zhang, L.; Cen, Y.; Wu, T.; Huang, C.; Baig, M.H.A. Comparison of the Continuity of Vegetation Indices Derived from Landsat 8 OLI and Landsat 7 ETM+ Data among Different Vegetation Types. *Remote Sens.* **2015**, *7*, 13485–13506, doi:10.3390/rs71013485.
56. Teillet, P.; Fedosejevs, G.; Barker, J.; Miskey, C.; Bannari, A. Spectral Simulations of Vegetation Indices in the Context of Landsat Data Continuity. In Proceedings of the 2006 IEEE International Symposium on Geoscience and Remote Sensing; IEEE, July 2006; pp. 1784–1787.
57. Teillet, P.M.; Ren, X. Spectral Band Difference Effects on Vegetation Indices Derived from Multiple Satellite Sensor Data. *Can. J. Remote Sens.* **2008**, *34*, 159–173.
58. Laliberte, A.S.; Goforth, M.A.; Steele, C.M.; Rango, A. Multispectral Remote Sensing from Unmanned Aircraft: Image Processing Workflows and Applications for Rangeland Environments. *Remote Sens.* **2011**, *3*, 2529–2551, doi:10.3390/rs3112529.
59. van Leeuwen, W.J.D.; Orr, B.J.; Marsh, S.E.; Herrmann, S.M. Multi-Sensor NDVI Data Continuity: Uncertainties and Implications for Vegetation Monitoring Applications. *Remote Sens. Environ.* **2006**, *100*, 67–81, doi:10.1016/j.rse.2005.10.002.
60. Baret, F.; Jacquemoud, S.; Guyot, G.; Leprieur, C. Modeled Analysis of the Biophysical Nature of Spectral Shifts and Comparison with Information Content of Broad Bands. *Remote Sens. Environ.* **1992**, *41*, 133–142, doi:10.1016/0034-4257(92)90073-S.
61. Jacquemoud, S.; Verhoef, W.; Baret, F.; Bacour, C.; Zarco-Tejada, P.J.; Asner, G.P.; François, C.; Ustin, S.L. PROSPECT+SAIL Models: A Review of Use for Vegetation Characterization. *Remote Sens. Environ.* **2009**, *113*, S56–S66, doi:10.1016/j.rse.2008.01.026.
62. Verrelst, J.; Camps-Valls, G.; Muñoz-Mari, J.; Rivera, J.P.; Veroustraete, F.; Clevers, J.G.P.W.; Moreno, J. Optical Remote Sensing and the Retrieval of Terrestrial Vegetation Bio-Geophysical Properties – A Review. *ISPRS J. Photogramm. Remote Sens.* **2015**, *108*, 273–290, doi:10.1016/j.isprsjprs.2015.05.005.
63. Magenau, E.; Clifton-Brown, J.; Awty-Carroll, D.; Ashman, C.; Ferrarini, A.; Kontek, M.; Martani, E.; Roderick, K.; Davey, C.; Amaducci, S.; et al. Site Impacts Nutrient Translocation Efficiency in Intra- and Interspecies *Miscanthus* Hybrids on Marginal Lands. *GCB Bioenergy* **2022**, submitted.
64. Tejera, M.D.; Miguez, F.E.; Heaton, E.A. The Older Plant Gets the Sun: Age-Related Changes in *Miscanthus* × *Giganteus* Phenology. *GCB Bioenergy* **2021**, *13*, 4–20, doi:10.1111/GCBB.12745.
65. Farrell, A.D.; Clifton-Brown, J.C.; Lewandowski, I.; Jones, M.B. Genotypic Variation in Cold Tolerance Influences the Yield

- of *Miscanthus*. *Ann. Appl. Biol.* **2006**, *149*, 337–345, doi:10.1111/j.1744-7348.2006.00099.x.
66. Datt, B. Remote Sensing of Water Content in Eucalyptus Leaves. *Aust. J. Bot.* **1999**, *47*, 909, doi:10.1071/BT98042.
  67. Miura, T.; Yoshioka, H.; Fujiwara, K.; Yamamoto, H. Inter-Comparison of ASTER and MODIS Surface Reflectance and Vegetation Index Products for Synergistic Applications to Natural Resource Monitoring. *Sensors* **2008**, *8*, 2480–2499, doi:10.3390/s8042480.
  68. Gitelson, A.A.; Kaufman, Y.J.; Merzlyak, M.N. Use of a Green Channel in Remote Sensing of Global Vegetation from EOS-MODIS. *Remote Sens. Environ.* **1996**, *58*, 289–298, doi:10.1016/S0034-4257(96)00072-7.
  69. Sripada, R.P.; Heiniger, R.W.; White, J.G.; Meijer, A.D. Aerial Color Infrared Photography for Determining Early In-Season Nitrogen Requirements in Corn. *Agron. J.* **2006**, *98*, 968–977, doi:10.2134/agronj2005.0200.
  70. Gitelson, A.A. Wide Dynamic Range Vegetation Index for Remote Quantification of Biophysical Characteristics of Vegetation. *J. Plant Physiol.* **2004**, *161*, 165–173, doi:10.1078/0176-1617-01176.
  71. Qi, J.; Chehbouni, A.; Huete, A.R.; Kerr, Y.H.; Sorooshian, S. A Modified Soil Adjusted Vegetation Index. *Remote Sens. Environ.* **1994**, *48*, 119–126, doi:10.1016/0034-4257(94)90134-1.
  72. Haboudane, D.; Miller, J.R.; Pattey, E.; Zarco-Tejada, P.J.; Strachan, I.B. Hyperspectral Vegetation Indices and Novel Algorithms for Predicting Green LAI of Crop Canopies: Modeling and Validation in the Context of Precision Agriculture. *Remote Sens. Environ.* **2004**, *90*, 337–352, doi:10.1016/j.rse.2003.12.013.
  73. Gitelson, A.; Merzlyak, M.N. Quantitative Estimation of Chlorophyll-a Using Reflectance Spectra: Experiments with Autumn Chestnut and Maple Leaves. *J. Photochem. Photobiol. B Biol.* **1994**, *22*, 247–252, doi:10.1016/1011-1344(93)06963-4.
  74. Rouse, J.W.; Haas, R.H.; Schell, J.A.; Deering, D.W. Monitoring Vegetation Systems in the Great Plains with ERTS. **1973**.
  75. Rondeaux, G.; Steven, M.; Baret, F. Optimization of Soil-Adjusted Vegetation Indices. *Remote Sens. Environ.* **1996**, *55*, 95–107, doi:10.1016/0034-4257(95)00186-7.
  76. Wu, C.; Niu, Z.; Tang, Q.; Huang, W. Estimating Chlorophyll Content from Hyperspectral Vegetation Indices: Modeling and Validation. *Agric. For. Meteorol.* **2008**, *148*, 1230–1241, doi:10.1016/j.agrformet.2008.03.005.
  77. Huete, A.. A Soil-Adjusted Vegetation Index (SAVI). *Remote Sens. Environ.* **1988**, *25*, 295–309, doi:10.1016/0034-4257(88)90106-X.
  78. Rusinowski, S.; Krzyżak, J.; Clifton-Brown, J.; Jensen, E.; Mos, M.; Webster, R.; Sitko, K.; Pogrzeba, M. New *Miscanthus* Hybrids Cultivated at a Polish Metal-Contaminated Site Demonstrate High Stomatal Regulation and Reduced Shoot Pb and Cd Concentrations. *Environ. Pollut.* **2019**, *252*, 1377–1387, doi:10.1016/j.envpol.2019.06.062.
  79. Urrego, J.P.F.; Huang, B.; Naess, J.S.; Hu, X.; Cherubini, F. Meta-Analysis of Leaf Area Index, Canopy Height and Root Depth of Three Bioenergy Crops and Their Effects on Land Surface Modeling. *Agric. For. Meteorol.* **2021**, *306*, 108444, doi:10.1016/j.agrformet.2021.108444.
  80. Lehnert, L.W.; Meyer, H.; Obermeier, W.A.; Silva, B.; Regeling, B.; Bendix, J. Hyperspectral Data Analysis in R : The Hsdr Package. *J. Stat. Softw.* **2019**, *89*, doi:10.18637/jss.v089.i12.
  81. Wood, S.N. *Generalized Additive Models*; Chapman and Hall/CRC, 2017; ISBN 9781315370279.
  82. Nolè, A.; Rita, A.; Ferrara, A.M.S.; Borghetti, M. Effects of a Large-Scale Late Spring Frost on a Beech (*Fagus Sylvatica* L.) Dominated Mediterranean Mountain Forest Derived from the Spatio-Temporal Variations of NDVI. *Ann. For. Sci.* **2018**, *75*, 83, doi:10.1007/s13595-018-0763-1.
  83. Antonucci, G.; Croci, M.; Miras-Moreno, B.; Fracasso, A.; Amaducci, S. Integration of Gas Exchange With Metabolomics: High-Throughput Phenotyping Methods for Screening Biostimulant-Elicited Beneficial Responses to Short-Term Water Deficit. *Front. Plant Sci.* **2021**, *12*, doi:10.3389/fpls.2021.678925.
  84. Kuhn, M. Building Predictive Models in R Using the Caret Package. *J. Stat. Softw.* **2008**, *28*, doi:10.18637/jss.v028.i05.
  85. Woźniak, G.; Dyderski, M.K.; Kompała-Bąba, A.; Jagodziński, A.M.; Pasierbiński, A.; Błońska, A.; Bierza, W.; Magurno, F.; Sierka, E.; Kompała-Bąba, A.; et al. Use of Remote Sensing to Track Postindustrial Vegetation Development. *L. Degrad. Dev.* **2021**, *32*, 1426–1439, doi:10.1002/ldr.3789.

86. Biecek, P. DALEX: MoDel Agnostic Language for Exploration and EXplanation 2018.
87. Li, P.; Jiang, L.; Feng, Z. Cross-Comparison of Vegetation Indices Derived from Landsat-7 Enhanced Thematic Mapper Plus (ETM+) and Landsat-8 Operational Land Imager (OLI) Sensors. *Remote Sens.* **2013**, *6*, 310–329, doi:10.3390/rs6010310.
88. Cui, Z.; Kerekes, J.P. Potential of Red Edge Spectral Bands in Future Landsat Satellites on Agroecosystem Canopy Green Leaf Area Index Retrieval. *Remote Sens.* **2018**, *10*, 1458, doi:10.3390/rs10091458.
89. Rengarajan, R.; Schott, J. Evaluation of Sensor and Environmental Factors Impacting the Use of Multiple Sensor Data for Time-Series Applications. *Remote Sens.* **2018**, *10*, 1678, doi:10.3390/rs10111678.
90. Kim, Y.; Huete, A.; Miura, T.; Jiang, Z. Spectral Compatibility of Vegetation Indices across Sensors: Band Decomposition Analysis with Hyperion Data. *J. Appl. Remote Sens.* **2010**, *4*, 043520, doi:10.1117/1.3400635.
91. Villaescusa-Nadal, J.L.; Franch, B.; Roger, J.-C.; Vermote, E.F.; Skakun, S.; Justice, C. Spectral Adjustment Model's Analysis and Application to Remote Sensing Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 961–972, doi:10.1109/JSTARS.2018.2890068.
92. Guo, Y.; Senthilnath, J.; Wu, W.; Zhang, X.; Zeng, Z.; Huang, H. Radiometric Calibration for Multispectral Camera of Different Imaging Conditions Mounted on a UAV Platform. *Sustainability* **2019**, *11*, 978, doi:10.3390/su11040978.
93. Guillen-Climent, M.L.; Zarco-Tejada, P.J.; Villalobos, F.J. Estimating Radiation Interception in Heterogeneous Orchards Using High Spatial Resolution Airborne Imagery. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 579–583, doi:10.1109/LGRS.2013.2284660.
94. Tao, H.; Feng, H.; Xu, L.; Miao, M.; Yang, G.; Yang, X.; Fan, L. Estimation of the Yield and Plant Height of Winter Wheat Using UAV-Based Hyperspectral Images. *Sensors* **2020**, *20*, 1231, doi:10.3390/s20041231.
95. Shah, S.H.; Angel, Y.; Houborg, R.; Ali, S.; McCabe, M.F. A Random Forest Machine Learning Approach for the Retrieval of Leaf Chlorophyll Content in Wheat. *Remote Sens.* **2019**, *11*, 920, doi:10.3390/rs11080920.
96. Tillack, A.; Clasen, A.; Kleinschmit, B.; Förster, M. Estimation of the Seasonal Leaf Area Index in an Alluvial Forest Using High-Resolution Satellite-Based Vegetation Indices. *Remote Sens. Environ.* **2014**, *141*, 52–63, doi:10.1016/j.rse.2013.10.018.
97. Volpato, L.; Pinto, F.; González-Pérez, L.; Thompson, I.G.; Borém, A.; Reynolds, M.; Gérard, B.; Molero, G.; Rodrigues, F.A. High Throughput Field Phenotyping for Plant Height Using UAV-Based RGB Imagery in Wheat Breeding Lines: Feasibility and Validation. *Front. Plant Sci.* **2021**, *12*, 185, doi:10.3389/fpls.2021.591587.
98. Prasad, N.R.; Patel, N.R.; Danodia, A. Cotton Yield Estimation Using Phenological Metrics Derived from Long-Term MODIS Data. *J. Indian Soc. Remote Sens.* **2021**, *49*, 2597–2610, doi:10.1007/s12524-021-01414-6.