UNIVERSITÀ CATTOLICA DEL SACRO CUORE

Sede di Piacenza

Dottorato di ricerca per il Sistema Agro-alimentare

Ph.D. in Agro-Food System

Cycle XXXIV

S.S.D. AGR/02

UAV multispectral remote sensing for high-throughput phenotyping of hemp and miscanthus traits

Candidate:

Giorgio Impollonia

Matriculation n: 4814778

Academic Year 2020/2021

UNIVERSITÀ CATTOLICA DEL SACRO CUORE

Sede di Piacenza

Dottorato di ricerca per il Sistema Agro-alimentare

Ph.D. in Agro-Food System

Cycle XXXIV

S.S.D. AGR/02

UAV multispectral remote sensing for high-throughput phenotyping of hemp and miscanthus traits

Coordinator:

Ch.mo Prof. Paolo Ajmone Marsan

Tutor: Prof. Stefano Amaducci

Candidate:

Giorgio Impollonia

Matriculation n: 4814778

Academic Year 2020/2021

Abstract

The demand for biomass to support bio-based production chains will increase over the next years, and this demand makes biomass production compete with food production on agricultural land. To avoid this competition, biomass crops can be grown on marginal land. However, their yield and quality on marginal land is highly variable because they are not genetically enhanced for biomass-related traits. Therefore, improving the biomass yield and quality of novel biomass crop genotypes through breeding programs is critical to the economic viability of growing them on marginal land. Examples of low-input biomass crops that can grow on marginal land and produce non-food biomass are miscanthus and hemp, two crops targeted in the EU project GRACE (GRowing Advanced industrial Crops on marginal lands for biorEfineries). The project aims to evaluate the sustainability of the increased biomass production of novel miscanthus and hemp genotypes on marginal lands, thus reducing the conflict with food production.

This thesis was carried out within the project's framework. Field trials were carried out to assess the performance of 14 miscanthus hybrids and 2 hemp cultivars through field measurements of relevant crops traits in different years and locations, in order to phenotype the novel genotypes. Unmanned aerial vehicles (UAVs) based multispectral remote sensing platform was used for high-throughput phenotyping (HTP) of hemp and miscanthus traits. UAV remote sensing acquires spectral reflectance signatures through multispectral sensors used to calculate vegetation indices (VIs) linked with crop traits. For hemp the crop traits estimated from UAV remote sensing were leaf area index (LAI) and leaf chlorophyll content (LCC); for miscanthus, light interception, plant height, green leaf biomass, standing biomass and moisture content. Three estimation methods were investigated: i) nonparametric regression methods, ii) physically based model inversion methods, and iii) hybrid regression methods. For hemp traits estimation, the PROSAIL model was used, and two inversion methods were compared: the look-up table (LUT) based on a cost function (physically based model inversion method) and the hybrid regression method based on machine learning (ML) algorithms. The hybrid regression methods performed better than LUT methods, both for LAI and LCC estimation, and the best accuracies were achieved by random forest (RF) for the LAI (0.76 m^2 m^{-2} of RMSE) and gaussian process regression (GPR) for the LCC (10.39 µg cm^{-2} of RMSE). For miscanthus traits estimation, the random forest ML algorithm (nonparametric regression method) was used, and it estimated with good accuracy light interception (8.4 % of RMSE), plant height (42 cm of RMSE), and moisture content (5.6 % of RMSE). Yield prediction, obtained using peak derived from the time series of VIs and random forest (RF) model, was performed to improve logistics biomass supply chain of miscanthus. The RF model accurately predicted miscanthus yield with 2.3

Mg DM ha^{-1} of RMSE. The RF model operability was evaluated through a timeline of the performance of the model using peak derived from partial VIs time series, and it showed a good capability to predict the yield three to seven months before the harvest. The HTP of hemp and miscanthus was carried out by applying the generalized additive model (GAM) to the time series of traits values (LAI and LCC for hemp and moisture content for miscanthus), which were estimated by the best estimation models from multiple multispectral UAV flights. The GAM analysis showed differences in the LAI and LCC dynamics between two hemp cultivars and differences in the moisture content dynamics between the novel miscanthus hybrids. Furthermore, to overcome the issue of UAV multi-sensor interoperability, linking equations derived from the PROSAIL model were used to link VIs of two different UAV multispectral sensors.

In conclusion, this linking procedure is interesting for crop phenotyping, where the field trials are often carried out in multi-location and with different UAV sensors. ML and inversion methods of the PROSAIL model estimated the crop traits with good accuracy, demonstrating that the UAV multispectral remote sensing platform is a suitable tool for HTP. UAV remote sensing enabled crop phenotypic traits to be distinguished and deepened our understanding of the traits' dynamics in contrasting genotypes throughout the growing season. Combining estimation models and GAM modelling applied to time series of crop trait values estimated from multiple multispectral images of UAV flights proved to be a powerful tool for HTP.

Keywords: UAV, remote sensing, high-throughput phenotyping, hemp, miscanthus, trait estimation, yield prediction, PROSAIL, machine learning, GAM, multi-sensor interoperability.

Contents

Chapter 1

Introduction

Introduction

Abstract

This chapter firstly provides a brief introduction to biomass crops on marginal lands, to the unmanned aerial vehicle (UAV) remote sensing for crop phenotyping and to the statistical and physical methods for estimating crop traits. Thereafter it outlines the objectives and content of the thesis.

Contents

1.1 Biomass crops on marginal land: role of miscanthus and hemp within EU GRACE project

Projections for the demand for biomass, to sustain bio-based production chains, will considerably increase in the next 30 years and dedicated field crops will have a prominent position in supporting this demand (Piotrowski *et al.*, 2015). This demand puts biomass production in competition with food production on agricultural land (Dauber *et al.*, 2012; Gelfand *et al.*, 2013). Dedicated crops for biomass production could therefore be grown on marginal lands to avoid competition with food production (Carlsson *et al.*, 2017; Fritsche *et al.*, 2010; Kang *et al.*, 2013; Tilman *et al.*, 2009). Marginal lands are areas not used for food production because of their low agronomic and economic potential (Schmidt *et al.*, 2015; Shortall, 2013), and are defined as marginal because of factors such as poor soil quality, contamination by heavy metals, poor water availability, steep slopes or distance from transportation (Pancaldi & Trindade, 2020). Biomass crops generally require low inputs and show high resource use efficiencies, so they grow well on marginal lands (Mehmood *et al.*, 2017). However, most biomass crops show very variable biomass yield and quality (Zegada-Lizarazu *et al.*, 2010) because they have not undergone genetic improvement for biomass-related traits (Jones *et al.*, 2015; Zhu *et al.*, 2016). Improving biomass yield and quality in novel biomass crop species for marginal lands is crucial to enable the economic viability of the cultivation of biomass crops on marginal lands (Pancaldi & Trindade, 2020). Biomass yield is an important trait, and its improvement is determined by other traits positively correlated with biomass yield, such as light interception, plant height and leaf characteristics (Boe & Beck, 2008; Fernandez *et al.*, 2009). Crop growth characteristics such as duration and timing affect biomass yield and can be maximized by breeding programs which aim to achieve crops with early leaf development, delayed plant senescence and late flowering time (Pancaldi & Trindade, 2020). As a result of climate change, the poor conditions of marginal lands will worsen because of abiotic stress intensification (Quinn *et al.*, 2015). The development of novel genotypes that couple robustness with optimal yields is fundamental for cultivating biomass crops on marginal lands (Jones *et al.*, 2015) and for the success of bio-based production chains. Novel miscanthus and hemp genotypes have recently been tested along with other biomass crops in two EU projects (OPTIMSC and MultiHemp) to evaluate their biomass production on marginal land (Lewandowski *et al.*, 2016; Tang *et al.*, 2017). Although both miscanthus and hemp are low-input crops (McCalmont *et al.*, 2017; Struik *et al.*, 2000) that can grow on marginal land and produce non-food biomass (Amaducci *et al.*, 2015), the EU projects OPTIMSC and MultiHemp demonstrated the importance of using genotypes suited to the environment and the intended end-use (Lewandowski *et al.*, 2016; Tang *et al.*, 2016). Genetic variations of miscanthus and hemp enable the

germplasm to be adapted to the specific needs of different industrial applications. The on-going EU project GRACE (GRowing Advanced industrial Crops on marginal lands for biorEfineries) aims to reduce the conflict with food production by unlocking the use of marginal lands for biomass production and demonstrate the sustainability of the increased biomass production of miscanthus and hemp on marginal lands. The GRACE project evaluates novel seed-based miscanthus hybrids from breeding programs at the Universities of Aberystwyth and Wageningen. The productive suitability of these miscanthus hybrids is evaluated in field trials conducted over a wide geographical and environmental range. These trials assess the performance of the miscanthus hybrids, as well as hemp cultivars, via field measurements of relevant crop traits, with the aim of phenotyping the novel genotypes. The phenotyping activities of the miscanthus and hemp traits in turn increase the knowledge of these novel genotypes and this information can be used to improve the crop germplasm in future breeding programs.

1.2 High-throughput phenotyping by UAV remote sensing: sensors and applications

Crop high-throughput phenotyping, which uses numerous sensors able to collect large amounts of data on genotype variability (Shi *et al.*, 2016) in fields, is considered a relevant scientific topic (Araus & Cairns, 2014; Tester & Langridge, 2010) for determining the phenotypes associated with novel genotypes (Furbank & Tester, 2011; Zaman-Allah *et al.*, 2015). Crop phenotypic traits are determined by genotype x environment interactions (Yang *et al.*, 2017) and can be divided into biophysical traits (e.g. biomass, height and leaf area index), biochemical traits (e.g. chlorophyll and nitrogen) and physiological traits (e.g. photosynthesis) (Xie & Yang, 2020). Rapid methods to characterize crop traits are essential in plant breeding (Banerjee *et al.*, 2020). However, traditional phenotyping methods for evaluating the phenotypic traits are time-consuming as the collection of phenotypic data is done manually (Araus & Cairns, 2014). This represents a bottleneck for the selection of phenotypic traits of novel genotypes (White *et al.*, 2012). Therefore, high-throughput phenotyping (HTP) platforms based on remote sensing, such as satellites and unmanned aerial vehicles (UAVs), have recently been used (Shi *et al.*, 2016; Wang *et al.*, 2019). For HTP applications, the satellite platform, due to the lack of spatial resolution for the identification of crop traits, is not suitable for monitoring small plots of field trials with numerous genotypes (Gevaert *et al.*, 2015; Han-Ya *et al.*, 2010). On the contrary, the UAV platform is a suitable tool for HTP (Berni *et al.*, 2009; Liebisch *et al.*, 2015) for its ability to capture high-resolution images, so it has been widely used in breeding programs (Gracia-Romero *et al.*, 2019; Ostos-Garrido *et al.*, 2019; Su *et al.*, 2019; Varela *et al.*, 2021; Yang *et al.*, 2017; Zhou *et al.*, 2019). UAV based remote sensing for HTP has been employed in many crops

(Feng *et al.*, 2021). UAV platforms can be equipped with different types of sensors, such as Red-Green-Blue (RGB), multispectral, hyperspectral, thermal cameras, and LIDAR (Xie & Yang, 2020) which are used to monitor the crop's phenotypic traits (Rahaman *et al.*, 2015; Zhang & Kovacs, 2012). RGB, multispectral, and hyperspectral UAV sensors based on spectral imaging acquire the reflectance of crops on different spectral bands (Yang *et al.*, 2017). RGB sensors are commonly used by UAVs in crop phenotyping due to their high spatial resolution and relatively low cost (compared with other sensors), but they have only three bands in the visible spectral region (Feng *et al.*, 2021). On the contrary, the hyperspectral sensors acquire images with hundreds and even thousands of continuous bands in the visible and near-infrared (NIR) spectral regions, but the cost is higher than RGB and multispectral sensors (Feng *et al.*, 2021). Multispectral sensors represent a compromise between the cost and the number of spectral bands compared to RGB (low cost and low number of spectral bands) and hyperspectral sensors (high cost and high number of spectral bands). The multispectral sensors have a small number of spectral bands (between 4 and 10) that are generally in the visible and NIR spectral regions (Guo *et al.*, 2021). They are widely used for UAV HTP for their spatial and spectral resolution and because are usually small, light and with an acceptable cost (Xie & Yang, 2020). For this reason, a wide variety of UAV multispectral sensors are available on the market with different spectral characteristics and the most common are reported in Table 1.1.

Table 1.1 Comparison of centre (centre wavelength in nm) and of FWHM (full width at half maximum in nm) of the most common UAV multispectral sensors available on the market.

Sensor	Blue		Green		Red		Red edge		Near-infrared	
	centre	FWHM	centre	FWHM	centre	FWHM	centre	FWHM	centre	FWHM
DJI P4	450	16	560	16	650	16	730	16	840	26
MicaSense RedEdge-MX	475	32	560	27	668	14	717	12	840	57
Parrot Sequoia	-	-	550	40	660	40	735	10	790	40
Sentera 6X	475	30	550	20	670	30	715	10	840	20
SlantRange 4P	470	100	550	100	650	40	710	20	850	100
Tetracam Micro-MCA4	490	10	550	10	680	10	-	-	800	10
YuSense MS600 PRO	450	35	555	25	660	20	720	10	840	35

The UAV multispectral sensors can calculate several vegetation indices (VIs) used in precision agriculture applications. The VIs are obtained by the combination of spectral bands, such as the Normalized Difference Vegetation Index (NDVI) (Rouse *et al.*, 1973), the Enhanced Vegetation

Index (EVI) (Huete *et al.*, 2002) and the Modified Chlorophyll Absorption in Reflectance Index (MCARI) (Daughtry *et al.*, 2000), and they are widely used to monitor crop traits. The spectral reflectance and the VIs are used to estimate or predict many crop traits using statistical and physical methods (see section 1.3). Several crop traits have been estimated using UAV-based remote sensing, such as plant height (Volpato *et al.*, 2021), fraction of absorbed photosynthetically active radiation (Lv *et al.*, 2021), canopy cover (Makanza *et al.*, 2018), aboveground biomass (Han *et al.*, 2019), yield (Johansen *et al.*, 2020; Wang *et al.*, 2019), leaf area index (Potgieter *et al.*, 2017), green leaf area index (Blancon *et al.*, 2019), leaf chlorophyll content (Wan, Zhang, *et al.*, 2021), leaf nitrogen content (Xu *et al.*, 2021), stay-green (Liedtke *et al.*, 2020) and senescence dynamics (Hassan *et al.*, 2018). Crop yield is considered one of the most important traits for phenotyping (Xie & Yang, 2020). The yield trait of biomass crops has also been investigated using UAV remote sensing for HTP, as reported by Li *et al.* (2020) for switchgrass and Wang *et al.* (2019) for perennial ryegrass. Several remote sensing studies have focused on yield prediction based on land surface phenology (LSP) (de Beurs & Henebry, 2005; Ji *et al.*, 2021; Meroni, d'Andrimont, *et al.*, 2021) which considers the development of crops using time series of VIs (Peng *et al.*, 2018; Sakamoto *et al.*, 2013). LSP aims to determine spatial and interannual variation in crop phenology using various descriptors derived from the time series of VIs and these descriptors are used to predict yield (Evans & Shen, 2021). LSP descriptors include the start of the season (SOS), the peak of the growing season, the stay-green duration (onset of senescence), the end of the season (EOS), and the growing season length (de Beurs & Henebry, 2010). 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 (Montazeaud *et al*., 2016) and EVI2 (Liu *et al*., 2019) for grain yield and the peak of GNDVI for biomass yields of perennial grass (Hamada *et al*., 2021).

1.3 Estimation of phenotypic crop traits: statistical and physical methods

Crop trait estimations using remote sensing are grouped into two main methods: the statistical method and the physical method (Baret & Buis, 2008). Both methods have expanded to integrate hybrid methods over the last years (Verrelst, Camps-Valls, *et al.*, 2015). This methodological expansion has divided the estimation methods into four categories (Verrelst *et al.*, 2019): i) parametric regression methods, ii) nonparametric regression methods, iii) physically based model inversion methods, and iv) hybrid regression methods (Figure 1.1). Parametric regression methods are widely used for their simplicity and ease of development (e.g. linear regression) and are applied in crop phenotyping studies to link VIs derived from multispectral remote sensing with crop traits (Machwitz *et al.*, 2021).

Contrary to parametric regression methods, nonparametric regression methods optimize the regression algorithm by learning phase based on training data and by including weights (coefficients) which are adjusted to minimize the estimation error (Verrelst *et al.*, 2019). Among nonparametric regression methods, nonlinear nonparametric methods, also called machine learning regression algorithms, are typically used for crop trait estimations. Machine learning (ML) regression algorithms are increasingly used in HTP as many VIs exhibit nonlinear relationships with their associated crop traits. Several ML algorithms were used for crop trait estimation, including artificial neural network (ANN), k-nearest neighbour (kNN), gaussian regression process (GPR) and random forest (RF). The ML algorithms are also used for the yield prediction of many crops (Jeong *et al.*, 2016; Marques Ramos *et al.*, 2020). RF is one of the most often used algorithms in many remote sensing analyses (Holloway & Mengersen, 2018). RF does not suffer from overfitting, can manage a high training size and has proven to be robust to outliers and noise (Belgiu & Drăguţ, 2016). For example, the RF algorithm was used to estimate crop traits, such as crop biomass (Han *et al.*, 2019) and yield (Johansen *et al.*, 2020), from UAV multispectral images. However, the main limits of the ML models are in the training set size (Millard & Richardson, 2015) and in the unreliability of predictions made beyond the range of values of the parameters present in the training set (Shah *et al.*, 2019). In addition, the ML models have demonstrated a limited transferability to different environments, cropping systems and growing seasons (Vuolo *et al.*, 2013). For these reasons, the transferability of ML models must be an important topic to consider in future studies on UAV applications in agricultural sciences (Johansen *et al.*, 2020). Physically based model inversion methods such as radiative transfer models (RTMs) are widely used to overcome crop trait estimation transferability problems and with respect to statistical methods (parametric and nonparametric regression methods) have the advantage that they minimize the reliance on *in situ* data (Atzberger *et al.*, 2015). One of the most popular RTMs is the PROSAIL model (Berger *et al.*, 2018), which simulates canopy reflectance by combining the leaf PROSPECT model (Jacquemoud & Baret, 1990) and the canopy SAIL model (Verhoef, 1984). Recently, studies have demonstrated that coupling UAV remote sensing with the PROSAIL model is a suitable approach for crop trait estimations for HTP (Jay *et al.*, 2017; Wan, Zhang, *et al.*, 2021; Wan, Zhu, *et al.*, 2021). The look-up table (LUT) strategy is used for the inversion of the PROSAIL model. The LUT strategy is based on the generation of simulated spectral reflectance for several combinations of canopy and leaf parameters and, through the application of a cost function, the spectral reflectance that most closely resembles the measured one is identified (Verrelst *et al.*, 2019). However, the several combinations of values of the parameters can generate similar canopy reflectance causing the ill-posed problem of the PROSAIL model inversion (Atzberger, 2004). Indeed, the use of multiple solutions (instead of the single best solution) (Darvishzadeh *et al.*, 2008;

Sehgal *et al.*, 2016) and the use of priori knowledge on the ranges of input parameters (Meroni *et al.*, 2004) have been evaluated to resolve the ill-posed problem. The hybrid regression methods combine the physically based methods with nonparametric regression methods. The nonparametric models trained with simulated reflectance values with the PROSAIL model are used as inversion strategies for crop trait estimations. The hybrid regression methods exploit the generic properties of physically based methods with the flexibility and computational efficiency of nonparametric regression methods (Verrelst, Rivera, *et al.*, 2015). The hybrid regression methods used in several studies are based on machine learning algorithms such as ANN (Atzberger, 2010), GPR (Verrelst *et al.*, 2016) and RF (Doktor *et al.*, 2014). Generally, both statistical and physical methods can estimate crop traits, each with advantages and disadvantages. This suggests that UAV remote sensing based on crop trait estimations using statistical or physical methods can be used as a tool for high-throughput phenotyping (HTP) and could substitute the field measurements used to evaluate the crop's phenotypic traits.

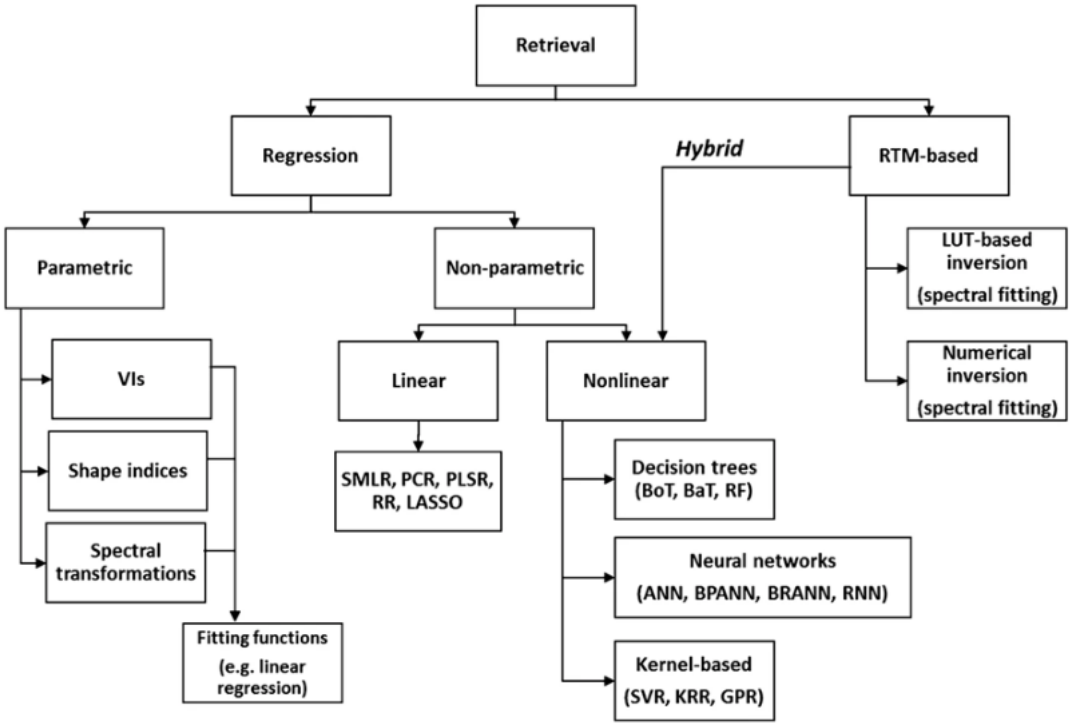


Figure 1.1 Schematic overview of the main estimation methods (Verrelst *et al.*, 2019).

1.4 Objectives

The use of suitable platforms for high-throughput screening of different hybrids of miscanthus and cultivars of hemp could support breeding programs. A suitable platform widely used in other crops

for high-throughput phenotyping (HTP) is the unmanned aerial vehicle (UAV). UAV remote sensing can phenotype crop traits through multispectral sensors that acquire spectral images useful for vegetation indices (VIs) calculation. However, the use of UAV multispectral remote sensing for HTP exhibit many research gaps which are:

G1. Issue of UAV multi-sensor interoperability.
G2. Use of the time series of VIs from UAV for yield prediction.
G3. Assessment of the models' transferability.
G4. Evaluation of the crop traits dynamics during the growing season using UAV remote sensing.
G5. Crop trait estimations of miscanthus and hemp for HTP.

Regarding the research gaps, many multispectral UAV sensors with different spectral characteristics and consequently with differences of VIs between sensors are available on the market, giving rise to the issue of UAV multi-sensor interoperability (research gap G1). Apart from this, the time series of VIs are widely used in satellite remote sensing applications for crop monitoring and for crop yield predictions using the descriptors of land surface phenology (LSP) derived from the time series of VIs. Indeed, the acquisition of data time series of VIs from UAV requires enough time to carry out multiple flights in the fields (research gap G2). However, the UAV market and the number of UAV users are constantly growing, and the use of the time series of VIs from UAV must be evaluated, in particular for HTP applications where the satellite platform is not suited to monitor small plots of field trials with numerous genotypes. Other research gaps of the HTP studies concern crop trait estimations. Indeed, statistical methods are typically used to link VIs to crop traits, but these models have a limited transferability to different environmental conditions (research gap G3). Therefore, new studies are needed to assess the transferability of statistical models, such as machine learning (ML), for UAV applications in HTP. The physical method, such as the PROSAIL model, is widely used to overcome the issue of transferability of crop trait estimations. However, to date, in the HTP applications, only time series VIs data from UAV and not the estimated crop traits of PROSAIL or ML models are used to evaluate the dynamics of the crop traits throughout the growing season (research gap G4). A statistical approach that can be an opportunity for the HTP from UAV remote sensing is the generalized additive model (GAM). GAM can capture the dynamic aspect of crop traits, and it has been successfully featured in several research fields. Therefore, research studies are needed to evaluate the combination of multiple UAV flights, estimation models and GAM analysis. Finally, there are no studies that have focused on the estimation of crop traits of hemp and miscanthus from UAV multispectral images for HTP applications (research gap G5). To better deepen the knowledge

on these gaps, the research question of this thesis is: "*Is UAV multispectral remote sensing a suitable platform for high-throughput phenotyping of hemp and miscanthus traits*?"

To answer this question, the objectives of the thesis are:

O1. To estimate hemp and miscanthus traits from UAV multispectral images for HTP.
O2. To evaluate the transferability of the ML models using independent datasets for model validation.
O3. To predict the yield for improving the logistics biomass supply chain of miscanthus.
O4. To evaluate the importance of VIs in the ML models for crop trait estimations and yield prediction of miscanthus.
O5. To identify the most accurate PROSAIL model inversion method for hemp trait estimations by comparing inversion methods.
O6. To link VIs calculated from UAV multispectral sensors with different spectral characteristics to improve multi-sensor interoperability.
O7. To phenotype the dynamics of the crop traits of contrasting miscanthus hybrids and hemp cultivars.

These objectives result in the following hypotheses:

H1. The ML models and PROSAIL model can be used to estimate traits of miscanthus hybrids and hemp cultivars.
H2. The quality training data can develop robust ML models to overcome the transferability problem.
H3. The peak derived from the VIs time series of UAV can be used to predict the yield of miscanthus.
H4. The spectral bands used for calculating the VIs have different importance depending on the crop traits to be estimated.
H5. The hybrid regression inversion methods will better estimate the crop traits than LUT inversion methods of the PROSAIL model.
H6. The PROSAIL model can be used to derive equations able to link multi-sensor VIs and to overcome the differences of VIs between sensors.
H7. The GAM analysis, applied to the time series of the crop traits estimated by ML or PROSAIL model inversion, can be used for phenotyping the dynamics of the crop traits of contrasting miscanthus hybrids and hemp cultivars.

1.5 Thesis outline

The chapters of this thesis are based on field research and the ones addressing hypotheses (1.4 Objectives) are reported in Table 1.2.

Table 1.2 Chapters and hypothesis.

Chapter	Hypothesis
2	H1, H5, H7
3	H1, H3, H4, H6
4	H1, H2, H4, H7

Chapter 2 compares hybrid and look-up table (LUT) inversion methods of the PROSAIL model for the leaf area index (LAI) and leaf chlorophyll content (LCC) estimation of hemp. It also explores the possibility of using GAM to combine multiple UAV flights and PROSAIL for phenotyping the LAI and LCC dynamics throughout the whole growing season. For this chapter, LAI and LCC measurements were collected on two contrasting hemp cultivars, green and yellow, under different nitrogen fertilisation levels during two growing seasons.

Chapter 3 estimates the crop traits, such as light interception, plant height, green leaf biomass and standing biomass, and predicts the yield of the novel miscanthus hybrids using the ML algorithm random forest from UAV multispectral images. This chapter also used the PROSAIL model to derive VIs link equations to improve sensor interoperability. The field trials were carried out in Italy and United Kingdom using two different UAV multispectral sensors.

Chapter 4 estimates the moisture content of miscanthus biomass from UAV multispectral images using the ML algorithm random forest (RF) and evaluates their transferability. This chapter also explores the application of GAM to moisture content time series estimated from the RF model for phenotyping the dynamics of senescence and identifying stay-green traits of contrasting miscanthus hybrids. The moisture content measurements were collected during two growing seasons in two locations.

Chapter 5 summarizes the main results of this thesis concerning the objectives and the hypotheses. This chapter also discusses the scientific contribution of this PhD thesis and provides an outlook on future research.

Chapter 2

Comparison of different inversion methods of the PROSAIL model for hemp trait estimations by UAV-based phenotyping

Comparison of different inversion methods of the PROSAIL model for hemp trait estimations by UAV-based phenotyping

Abstract

In this chapter, an unmanned aerial vehicle (UAV) multispectral remote sensing platform was used to estimate leaf area index (LAI) and leaf chlorophyll content (LCC) of two contrasting hemp cultivars, a green and a yellow one, during two growing seasons under four nitrogen fertilisation levels. The hemp traits were estimated by the inversion of the PROSAIL model from UAV multispectral data. Look-up table (LUT) and hybrid regression inversion methods were compared for LAI and LCC estimation. The LAI trait was estimated with more accuracy than the LCC trait. The hybrid methods performed better than LUT methods, both for LAI and LCC, and the best accuracies were achieved by random forest (RF) for the LAI (0.76 m^2 m^{-2} of RMSE) and by gaussian process regression (GPR) for the LCC (10.39 μg cm^{-2} of RMSE). High-throughput phenotyping (HTP) was carried out by applying the generalized additive model (GAM) to the time series of traits estimated by the best inversion methods of the PROSAIL model from multiple multispectral UAV flights. The GAM analysis showed differences in the LAI and LCC dynamics between two hemp cultivars and between the nitrogen fertilisation levels that significantly affected their dynamics. The HTP based on UAV remote sensing proved to be a powerful tool to estimate hemp traits and to improve our understanding of the traits' dynamics of contrasting cultivars throughout the whole growing season.

Keywords: Hemp, remote sensing, UAV, high-throughput phenotyping, trait estimation, PROSAIL, LUT, machine learning, inversion methods comparison, GAM.

Contents

2.1 Introduction

Interest in hemp *(Cannabis sativa* L.) cultivation, which is expanding internationally (Amaducci *et al.*, 2015), is related to the large number of products that can be obtained from its biomass (Crini & Lichtfouse, 2020), its high yield potential (up to 20 Mg ha^{-1}) (Burczyk *et al.*, 2008; Struik *et al.*, 2000; Tang *et al.*, 2016), low input requirements (Tang *et al.*, 2017), and positive environmental impact (van der Werf, 2004). All these features make hemp an ideal crop for the bio-based sector. In fact, not only hemp produces raw material for a wide range of bio-based applications, but it also produces seeds for food applications (Amaducci *et al.*, 2015; Tang *et al.*, 2016), in this way avoiding issues linked to indirect land use change. Despite the large interest in hemp, its cultivation has not significantly expanded in the last years due to legal problems related to the presence of psychoactive cannabinoids but also to the limited innovation along the whole value chain (Venturi *et al.*, 2007). In a recent EU project (MultiHemp), for the first time, innovative biotechnological tools have been applied for hemp breeding but so far the management of hemp cultivation and the monitoring of its growth have not benefitted from the application of innovative precision agriculture technologies.

Remote sensing is a precision agriculture technology widely used to monitor crop growth (de Castro *et al.*, 2021; Sishodia *et al.*, 2020). Remote sensing platforms, such as satellites and unmanned aerial vehicles (UAVs), acquire a large volume of spectral data with high spatial and temporal resolutions, which are needed for applications in both precision agriculture (Segarra *et al.*, 2020) and high-throughput phenotyping (HTP) in the frame of breeding programs (Guo *et al.*, 2021; Yang *et al.*, 2017). The spectral data acquired from the remote sensing HTP platforms are used to estimate crop traits during the growing season (Blancon *et al.*, 2019; Impollonia *et al.*, 2022). Leaf area index (LAI) and leaf chlorophyll content (LCC) are among the most important crop traits estimated in HTP applications (Potgieter *et al.*, 2017; Xie & Yang, 2020). Spatial and temporal information on LAI and LCC are usually regarded as relevant traits to monitor the status of crop growth (Duan *et al.*, 2019). Two main methods are used to estimate LAI and LCC using spectral data via remote sensing: i) statistical methods such as linear regression or machine learning, and ii) physical method based on radiative transfer models (RTMs) inversion. RTMs have the advantage, over the statistical method, to minimize the reliance on *in situ* data (Atzberger *et al.*, 2015). One of the most popular RTMs is the PROSAIL model (Berger *et al.*, 2018), which simulates the canopy reflectance by combining the leaf PROSPECT model (Jacquemoud & Baret, 1990) and the canopy SAIL model (Verhoef, 1984). Two main methods are commonly used for the PROSAIL model inversion: i) look-up tables (LUTs) (Atzberger *et al.*, 2015; Verrelst *et al.*, 2014; Verrelst, Rivera, *et al.*, 2015) based on a cost function; and ii) hybrid regression method (Verrelst *et al.*, 2019) based on machine learning techniques such

as artificial neural network (ANN) (Atzberger, 2010), gaussian process regression (GPR) (Verrelst *et al.*, 2016) and random forest (RF) (Doktor *et al.*, 2014). However, the diverse combinations of crop traits can generate similar canopy reflectance causing the ill-posed problem of the PROSAIL model inversion (Atzberger, 2004). To resolve this issue, several strategies have been evaluated, such as the use of multiple solutions (instead of the single best solution) (Darvishzadeh *et al.*, 2008; Sehgal *et al.*, 2016) and the use of a priori knowledge on the ranges of input parameters (Meroni *et al.*, 2004). Recently, several studies have demonstrated that the coupling of UAV remote sensing and the PROSAIL model enables reliable estimations of crop traits for HTP purposes, such as LAI and LCC (Duan *et al.*, 2014; Jay *et al.*, 2017; Sun *et al.*, 2021; Wan, Zhang, *et al.*, 2021; Zhu *et al.*, 2019). These studies only evaluated the ability of the PROSAIL model to estimate crop traits, without characterizing the dynamic of crop traits evolution along the growing season using the values estimated by the PROSAIL model. On the contrary, Impollonia *et al.* (2022) demonstrated that HTP obtained by combining multiple UAV flights, machine learning estimation models, and generalized additive model (GAM) can characterise the dynamics of crop traits. However, no studies have yet investigated this combination featuring PROSAIL instead of machine learning models to characterise the seasonal dynamic of crop traits for HTP. Additionally, to our knowledge, there is no study that has focused on the estimation of LAI and LCC of hemp from UAV multispectral images for precision agriculture applications.

In this context, the goals of this study were: i) to estimate the hemp traits, leaf area index (LAI) and leaf chlorophyll content (LCC), by UAV-based phenotyping using the inversion of the PROSAIL model, ii) to compare two inversion methods of the PROSAIL model: look-up table (LUT) and hybrid regression methods, and iii) to characterise the dynamics of crop traits (LAI and LCC) of two contrasting hemp cultivars (a yellow and a green cultivar) under different nitrogen fertilisation levels via generalized additive model (GAM).

2.2 Materials and methods

2.2.1 Experimental design

The field experiments were conducted at the CERZOO research centre (45°00′11.70″ N, 9°42′35.39″ E) in the province of Piacenza (NW Italy) during the years 2020 and 2021 (Figure 2.1). The two cultivars 'Futura 75', a conventional green one, and 'Fibror 79', a yellow-stalked one, both developed and provided by Hemp-it (France), were sown on the 6th and on the 9th of April, in 2020 and 2021, respectively. Sowing density and depth were about 45 kg ha^{-1} and 3 cm, respectively. The experimental layout was a complete randomised block design with 4 levels of nitrogen fertilisation:

0, 25, 50 and 100 kg N ha^{-1}, with *n*=4 replicates for a total of *n*= 32 plots (see Blandinières (2022) for details on experimental design).

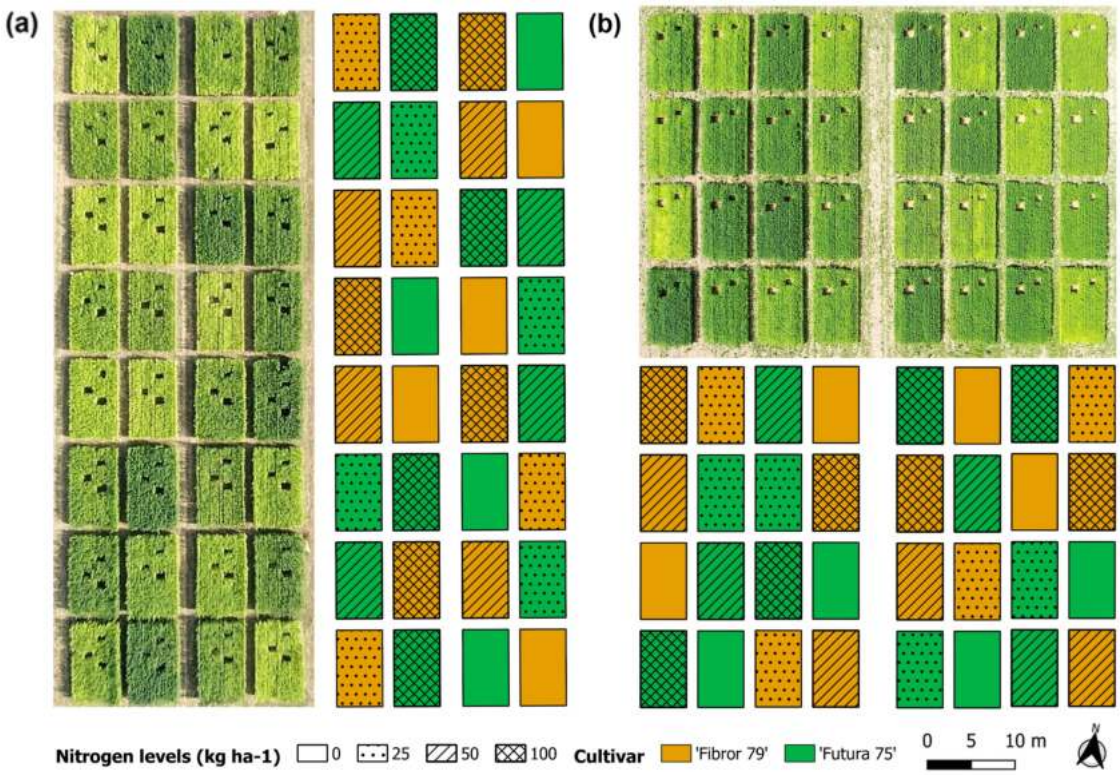


Figure 2.1 Experimental field design and drone picture of hemp trials in (a) 2020 and (b) 2021. The destructive samplings of 1 m^2 are observable in drone pictures.

2.2.2 Crop measurements

Four georeferenced samplings were carried out for each plot across the two growing seasons (total *n*=256) to measure leaf area index (LAI) (Figure 2.2). A ceptometer was used (AccuPAR LP-80, Decagon Devices, Inc., Pullman, Washington, USA) when the sun reached the zenith (between 12 a.m. and 2 p.m.) and five-six measurements per plot were taken to measure LAI non-destructively, on the same area that was sampled by Blandinières (2022) (Figure 2.1). At the same time (Figure 2.2), leaf dry content per unit leaf area (C_m) and leaf water per unit leaf area (C_w) were determined from all the leaves of a three-to-five plant subsample. The leaves were separated from the stems and transferred to a fridge at -18 °C. The leaf surface was determined by scanning the leaves, and the leaves were subsequently oven dried at 65 °C and weighted (see Blandinières (2022) for details on crop measurements). The C_m (g cm^{-2}) and C_w (g cm^{-2}) were calculated as the ratio of the dry weight (C_m) or of the water weight (C_w) of the leaves to their surface. Subsequently, leaf chlorophyll content

(LCC) samples were collected at the top of the canopy, rapidly stored in ice, and then stored in a fridge at -18 °C. Chlorophyll content was measured by using a chlorophyll analysis protocol based on Ritchie (2006) and Warren (2008) protocols. The chlorophyll analysis was only carried out on the samples from 2020.

2.2.3 UAV multispectral data

The unmanned aerial vehicle (UAV) DJI Matrice 210 RTK (SZ DJI Technology Co., Shenzhen, Guangzhou, China) was used in the experiment. The UAV was equipped with a MicaSense RedEdge-Mx (MicaSense, Seattle, WA, USA) camera to collect the multispectral images. RedEdge-Mx camera acquired the images in 5 spectral bands: blue (475 nm centre, 32 nm FWHM), green (560 nm centre, 27 nm FWHM), red (668 nm centre, 14 nm FWHM), red edge (717 nm centre, 12 nm FWHM) and near-infrared (840 nm centre, 57 nm FWHM). The UAV flights were carried out at the same time as crop measurements, and supplementary flight missions (Figure 2.2) were also carried out to improve the analysis of the crop traits dynamics derived from multiple UAV flights as suggested by Impollonia *et al.* (2022). All flights were performed in clear sky conditions between 11 a.m. and 3 p.m. The flight altitude above ground level (AGL) was 50 m, the forward overlap was set at 80% and the lateral overlap was set at 75% of the images. The ground sampling distance (GSD) was 2.78 cm, and the flight speed was set at 3 m s^{-1}. The DJI Pilot software (SZ DJI Technology Co., Shenzhen, Guangzhou, China) was used for the flight planning and automatic mission control. The reflectance panel provided by MicaSense, and the light sensor mounted at the top of the UAV were used for the radiometric calibration of the images. Pix4D mapper (Pix4D, S.A., Lausanne, Switzerland) software was used for radiometric calibration and orthomosaic generation. The experimental designs were drafted in AutoCAD (Autodesk, San Rafael, California, USA) and subsequently georeferenced using QGIS software (QGIS Development Team, 2021). Polygons of 1 m^2 were also drafted and georeferenced on the position of the destructive samplings, in order to extract the mean value of spectral data of each plot for the inversion PROSAIL model validation. For the time series analysis based on multiple UAV flights, the sampled quadrats were subtracted from the experimental designs to eliminate the noise caused by these destructive measurements on the multispectral images as shown in Figure 2.1.

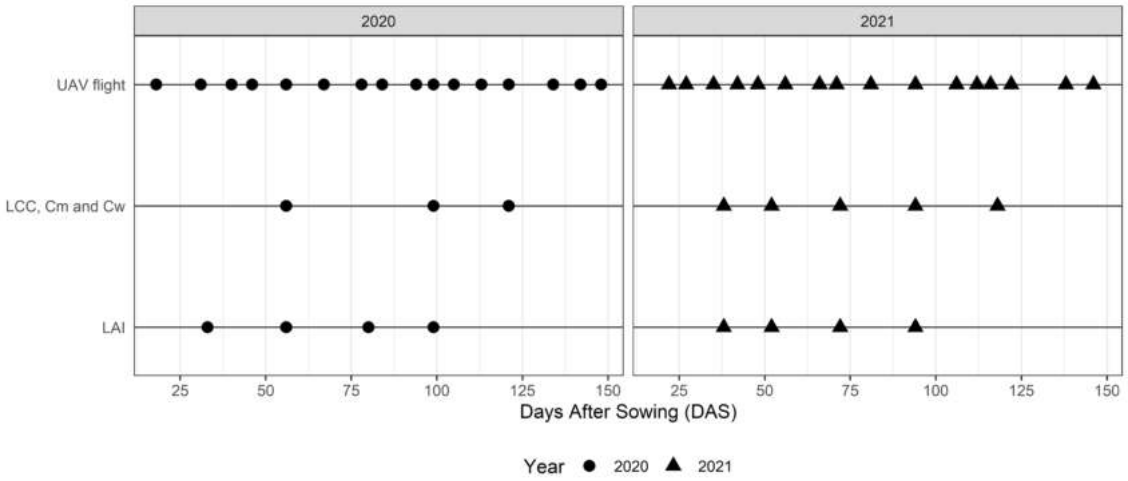


Figure 2.2 Seasonal calendar of UAV flights and crop measurements (LAI, LCC, Cm and Cw) during the two years (2020 and 2021).

2.2.4 PROSAIL model

LAI and LCC of hemp cultivars were estimated by the inversion of the PROSAIL model. The PROSAIL model combine the PROSPECT and SAIL models, simulating the canopy reflectance from 400 to 2500 nm, using 10 parameters of input. Four of these ten parameters were used to simulate the leaves' optical properties (PROSPECT model): leaf structure parameter (N), leaf chlorophyll content (LCC), leaf equivalent water thickness (C_w) and leaf dry matter content (C_m). The six other parameters were used to simulate the bidirectional reflectance of the canopy (SAIL model): leaf area index (LAI), average leaf inclination angle (ALIA), hotspot parameter (hot), solar zenith angle (tts), observer zenith angle (tto) and relative azimuth angle (psi). The soil reflectance was also considered in the PROSAIL model and was selected from the database 'ICRAF-ISRIC Soil VNIR Spectral Library' of the Soil Information System (ISIS) of the International Soil Reference and Information Centre (ISRIC). The choice of the soil reflectance was done, firstly by resampling the Italian soils reflectance based on UAV camera characteristics (MicaSense RedEdge-Mx), secondly by calculating the differences of soils reflectance between the Italian soils (soil database) and the soil observed in this study (the reflectance of soil was extracted from the UAV multispectral images) and finally the soil with less difference in reflectance was used in the PROSAIL model. In order to select the ranges of the canopy and leaf parameters, the field measurements acquired during the two growing seasons were used (see 2.2.2 Crop measurements), which is also an efficient way of reducing the ill-posed problem as suggested by Meroni *et al.* (2004). The ALIA range was set between 10 and 30 for the planophile nature of the hemp canopy (Meijer *et al.*, 1995). The PROSAIL inputs (parameter combinations) and outputs (spectral reflectance) are used for look-up table (LUT) generation. The

hsdar R package (Lehnert et *al.*, 2019) 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 LUT generated included 86,400 parameter combinations following the ranges (minimum and maximum) and the steps of the parameters summarized in Table 2.1. Regarding the parameter combinations used for LUT generation, all parameter combinations were considered for LAI and LCC estimation as the hemp cultivars evaluated in this study exhibited large differences in traits (e.g. high LAI and low LCC for 'Fibror 79' and high LAI and high LCC for 'Futura 75' at the end of the vegetative phase) throughout the whole growing season (e.g. low LAI and high LCC at the start of the growing season and low LAI and low LCC at the end of the growing season). The spectral reflectance simulated (outputs) were resampled based on MicaSense RedEdge-Mx characteristics camera (see section 2.2.3).

Table 2.1 Ranges of input parameters for the PROSAIL model for generating the LUT.

	Parameter	Abbreviation	Unit	Values
Leaf	Structure parameter	N	Unitless	1.5
	Chlorophyll content	LCC	µg cm^{-2}	5 – 60 (step =5)
	Equivalent water thickness	C_w	g cm^{-2}	0.006 – 0.03 (step = 0.004)
	Dry matter content	C_m	g cm^{-2}	0.004 – 0.007 (step = 0.001)
Canopy	Leaf area index	LAI	m^2 m^{-2}	0.1 – 6 (step = 0.3)
	Average leaf inclination angle	ALIA	deg	10 – 30 (step = 10)
	Hotspot parameter	hot	m m^{-1}	0.1
	Solar zenith angle	tts	deg	20 – 30 (step = 5)
	Observer zenith angle	tto	deg	10
	Relative azimuth angle	psi	deg	190 – 195 (step = 5)

2.2.5 Inversion methods of the PROSAIL model

Two inversion methods were compared in this study: look-up table method based on a cost function and hybrid regression method based on machine learning techniques.

2.2.5.1 The look-up table inversion method

The look-up table (LUT) was sorted using the cost function based on root mean square error (RMSE) to find the solution to the inverse problem for the measured canopy reflectance (Darvishzadeh *et al.*, 2008; Sehgal *et al.*, 2016). The $RMSE_r$ cost function (Equation 2.1), between the measured reflectance and the simulated reflectance found in the LUT, was calculated as:

$$RMSE_r = \sqrt{\frac{\sum_{i=1}^{n}\left(R_{measured_\lambda} - R_{simulated_\lambda}\right)^2}{n}} \qquad \text{Equation 2.1}$$

where n is the number of spectral bands, $R_{measured}$ is a measured reflectance at spectral band λ and $R_{simulated}$ is a simulated reflectance at spectral band λ in the LUT. Two LUT methods were tested to find the solution to the inversion problem. The first solution, single best solution, was found as the set of input parameters corresponding to the reflectance in the LUT that provides the smallest $RMSE_r$, and it was called here as LUT-I. However, this solution is not always the optimal solution since it may not be unique (ill-posed problem). In order to solve this problem, the second solution, multiple best solutions, was tested using the mean value of parameters corresponding to the best 100 solutions (i.e. having the smallest sorted $RMSE_r$) and it was called here as LUT-II.

2.2.5.2 The hybrid regression inversion method

The hybrid methods utilised the parameter combinations (*y*) and the simulated spectral reflectance (*x*) from the PROSAIL model, used for the LUT generation, for training a machine learning regression model. Therefore, the hybrid regression methods allow replacing the field measurements needed to train nonparametric models with PROSAIL input variables (Verrelst *et al.*, 2019). This study evaluated different machine learning regression models: random forest (RF), gaussian process regression (GPR), artificial neural network (ANN) and ensemble method (EM) obtained combining RF, GPR and ANN via stacking. The machine learning regression models were built using the *caret* and *caretEnsemble* R packages (Kuhn, 2008; Mayer, 2019). The structural hyperparameters of the machine learning regression models were optimized by grid-searching method using cross-validation. The training dataset was created using a stratified random sampling method by LAI, LCC and C_m values distribution of the LUT. The function of *caretList* was used for building the machine learning regression models using the method *rf*, *gaussprRadial* and *nnet*, for RF, GPR and ANN, respectively. The EM model was built using the function *caretStack* that finds a good linear combination of the models (RF, GPR and ANN).

2.2.5.3 Comparison of inversion methods

The field measurements of LAI and LCC were used to validate the inversion methods of the PROSAIL model. The root mean square error (RMSE, Equation 2.2) and the normalized root mean square error (NRMSE, Equation 2.3) were used for inversion methods comparison and was calculated for each method as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_i - y_i)^2}{n}} \quad \text{Equation 2.2}$$

$$NRMSE\ (\%) = \frac{\sqrt{\frac{\sum_{i=1}^{n}(x_i - y_i)^2}{n}}}{\bar{y}} 100 \quad \text{Equation 2.3}$$

where n is the sample number, x_i and y_i are the estimated and measured value of each trait, and $\bar{y}$ is the mean of the measured value. The performance metrics were also calculated for each season, for each cultivar, for each trait and for three different intervals of the traits. The LAI trait intervals investigated were less than 2 m^2 m^{-2}, between 2 m^2 m^{-2} and 4 m^2 m^{-2} and greater than 4 m^2 m^{-2}. The LCC trait intervals investigated were less than 20 µg cm^{-2}, between 20 µg cm^{-2} and 40 µg cm^{-2} and greater than 40 µg cm^{-2}. The LCC trait estimation was done only to the 2020 season because the LCC samples of 2021 were not analysed yet.

2.2.6 GAM for crop phenotyping

The hemp cultivar traits were estimated from multiple UAV flights (supplementary flights were also considered) using the best inversion methods for each trait for phenotyping the dynamics of LAI and LCC and identifying differences among cultivars and nitrogen fertilisation levels. The time series of LAI and LCC values estimated from the PROSAIL model inversion were fitted against the day after sowing (DAS). The statistical analysis of the hemp traits time series was carried out via a generalized additive model (GAM). The GAM is a non-parametric regression model which allows the integration of non-parametric smoothing functions and non-linear fitting of the variables. GAM models were fitted in R package *mgcv* (Wood, 2017). The fitted model used fixed factors such as season, block, cultivar, and nitrogen fertilisation levels, and a smooth for DAS, based on season and on interaction of cultivars and nitrogen fertilisation levels.

2.3 Results

2.3.1 Comparison of inversion methods for LAI trait estimation

The results of the comparison of the different methods used for the inversion of the PROSAIL model for leaf area index (LAI) estimation is reported in Figure 2.3. Generally, the hybrid methods achieved better accuracies than look-up table (LUT) methods. Random forest (RF) achieved the highest accuracy with 0.76 m^2 m^{-2} of RMSE and 26.8 % of NRMSE. LUT-I showed greater accuracy than LUT-II which ranked last. The hybrid method with the worst accuracy was ensemble method (EM) with 0.88 m^2 m^{-2} of RMSE and 31.3 % of NRMSE (Figure 2.3).

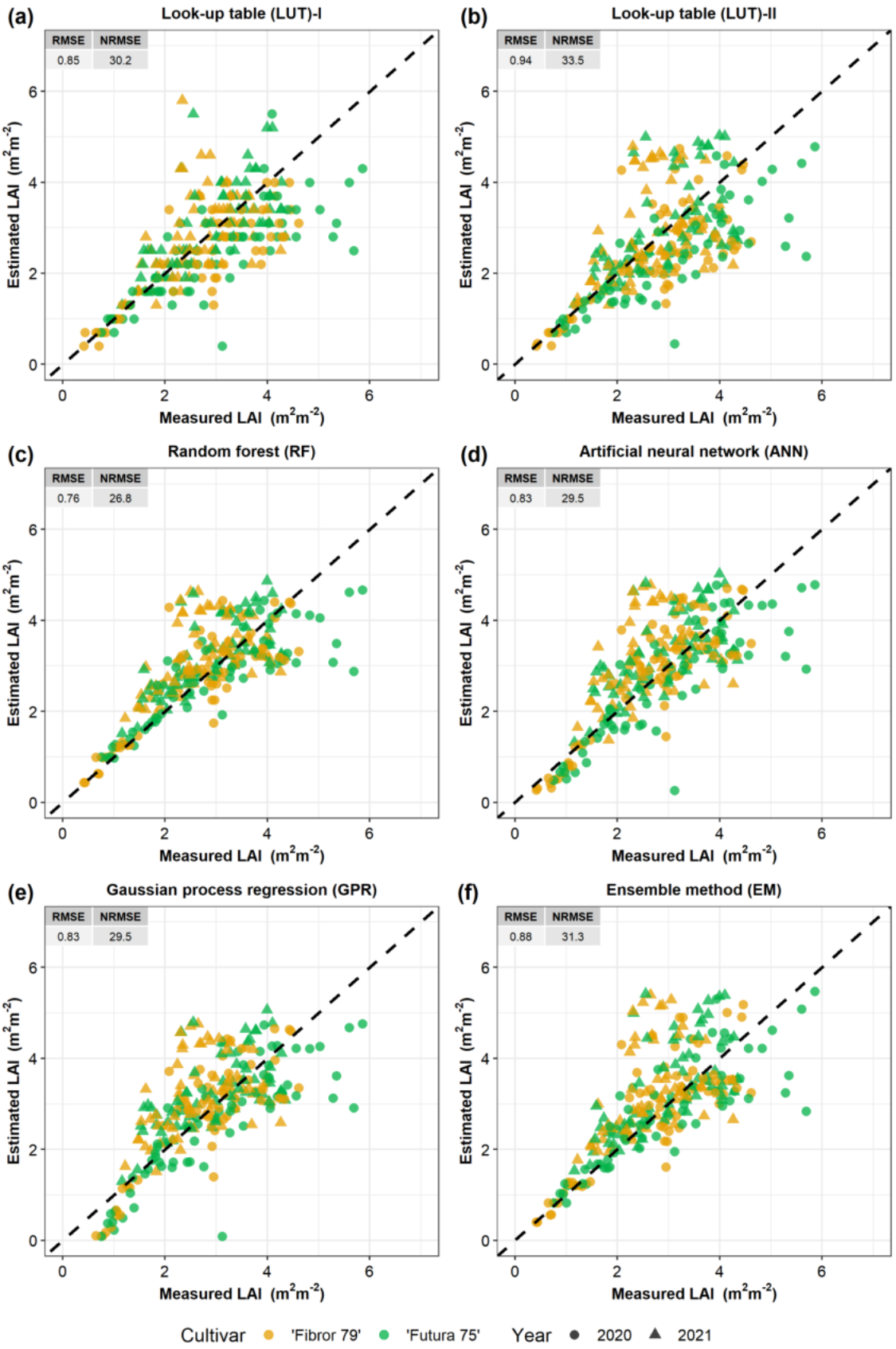


Figure 2.3 Estimated vs measured LAI of hemp with different inversion methods: (a) LUT-I, (b) LUT-II, (c) RF, (d) ANN, (e) GPR and (f) EM.

In Figure 2.4 the NRMSE values of the different methods, divided by years and cultivars, are reported. Generally, RF was the best method for both years and cultivars. The LUT methods showed high differences in NRMSE between years. 'Fibror 79' displayed a higher variability of NRMSE than 'Futura 75' across the two years of experimentations (Figure 2.4). The different inversion methods were evaluated on three LAI intervals (Figure 2.5). The LUT methods achieved the lowest NRMSE when LAI was less than 2 m^2 m^{-2} and achieved the highest NRMSE when LAI was greater than 4 m^2 m^{-2}. In the LAI interval ranging from 2 m^2 m^{-2} to 4 m^2 m^{-2}, RF achieved the best accuracy, but no relevant differences were observed between the other methods. The inversion methods, when LAI was less than 4 m^2 m^{-2} were more accurate for 'Futura 75' than 'Fibror 79' while the opposite occurred when LAI was greater than 4 m^2 m^{-2} (Figure 2.5).

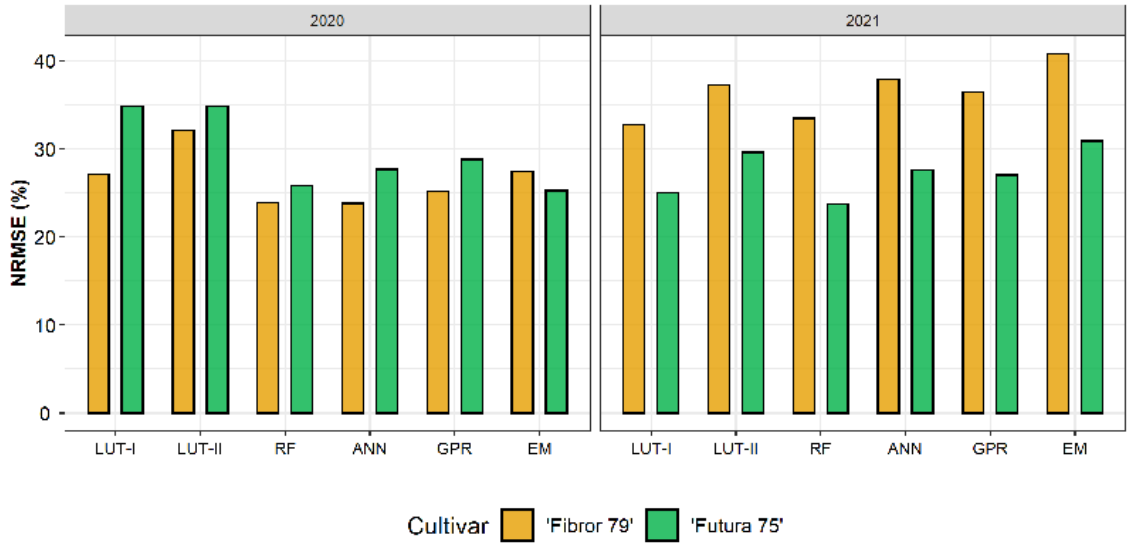


Figure 2.4 NRMSE of the inversion methods for LAI estimation according to the different cultivars and seasons.

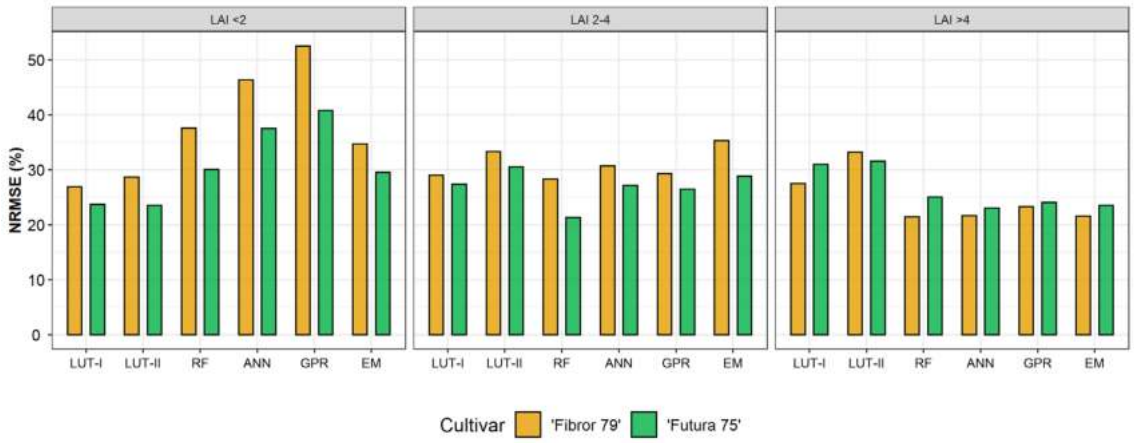


Figure 2.5 NRMSE of the inversion methods for LAI estimation according to the different cultivars and three LAI intervals (LAI < 2 m^2 m^{-2}, LAI 2-4 m^2 m^{-2} and LAI > 4 m^2 m^{-2}).

2.3.2 Comparison of inversion methods for LCC trait estimation

The results of the inversion methods of the PROSAIL model for the estimation of leaf chlorophyll content (LCC) are presented in Figure 2.6. The hybrid methods achieved better accuracies than LUT methods. Gaussian process regression (GPR) achieved the greatest accuracy with 10.39 µg cm^{-2} of RMSE and 40.5 % of NRMSE. The hybrid method that showed the worst accuracy was RF with 11.26 µg cm^{-2} of RMSE and 43.9 % of NRMSE (Figure 2.6). NRMSE values of the different methods and for both cultivars are reported in Figure 2.7. Overall, the two cultivars displayed a similar NRMSE across the different inversion methods (Figure 2.7). The different methods were evaluated on three LCC intervals (Figure 2.8). The inversion methods, when LCC was less than 20 µg cm^{-2} were more accurate for 'Fibror 79' than 'Futura 75' while the opposite occurred when LCC was greater than 40 µg cm^{-2}.

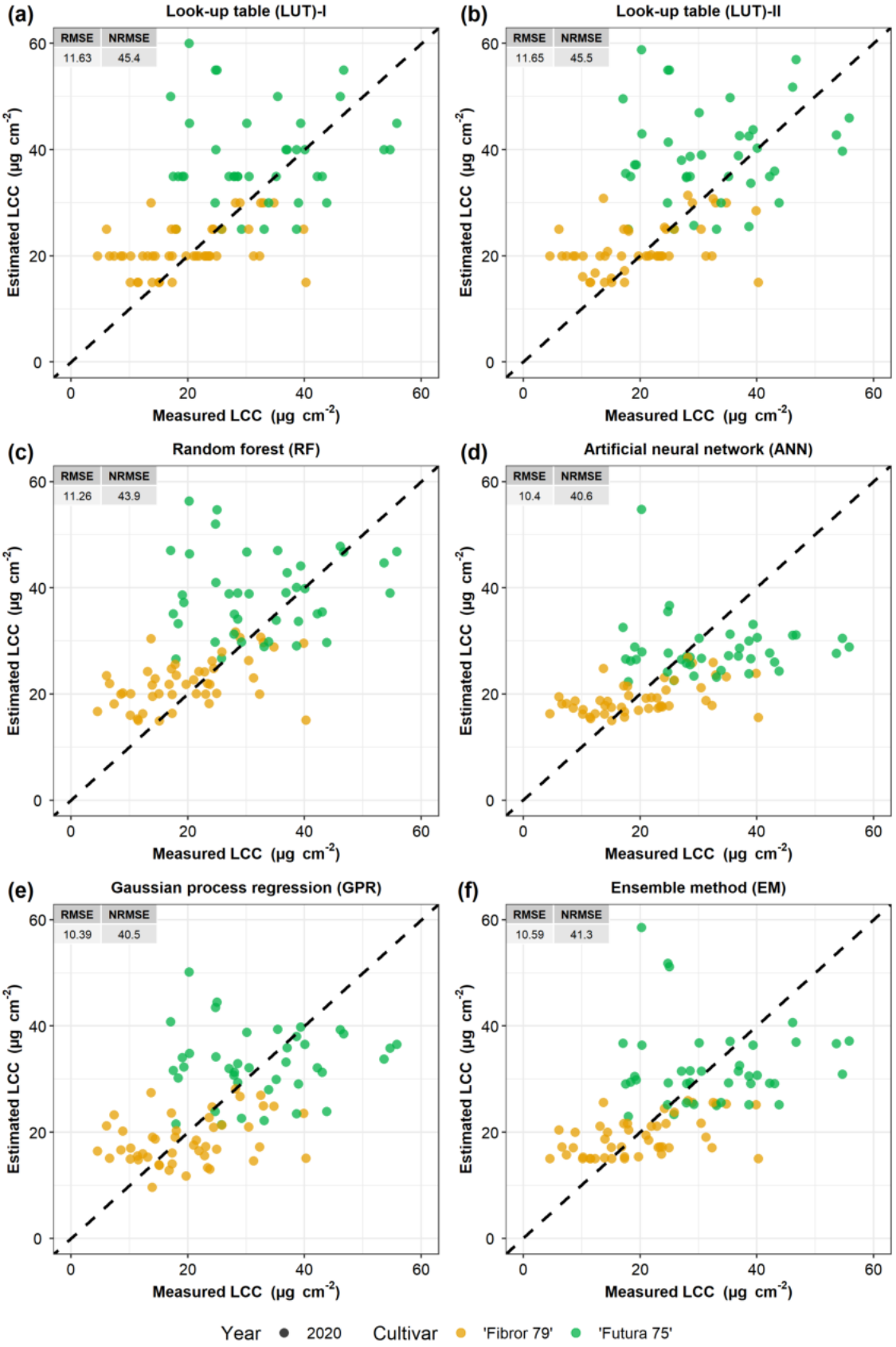


Figure 2.6 Estimated vs measured LCC of hemp with different inversion methods: (a) LUT-I, (b) LUT-II, (c) RF, (d) ANN, (e) GPR and (f) EM.

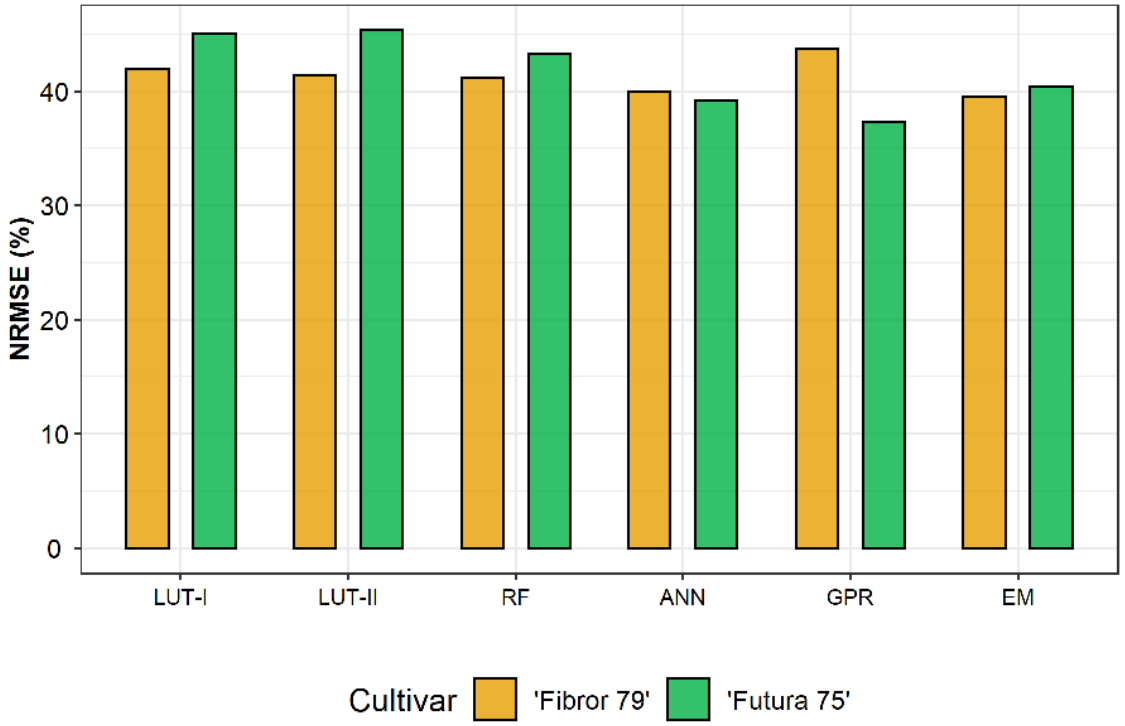


Figure 2.7 NRMSE of the inversion methods for LCC estimation according to the different cultivars.

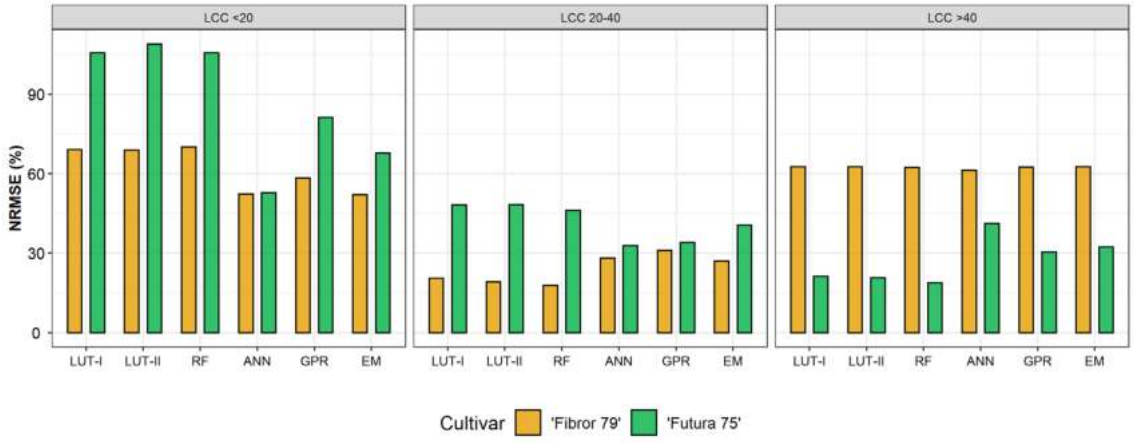


Figure 2.8 NRMSE of the inversion methods for LCC estimation according to the different cultivars and three LCC intervals (LCC < 20 µg cm^{-2}, LCC 20-40 µg cm^{-2} and LCC > 40 µg cm^{-2}).

2.3.3 Dynamics of LAI and LCC

The best inversion methods, RF for LAI and GPR for LCC, were used to estimate LAI and LCC of the two hemp cultivars using the spectral band acquired from multiple unmanned aerial vehicle (UAV) flights during 2020 and 2021 growing seasons. Maps of LAI and LCC estimated on 105 days after sowing (DAS) during the growing season 2020 are reported in Figure 2.9.

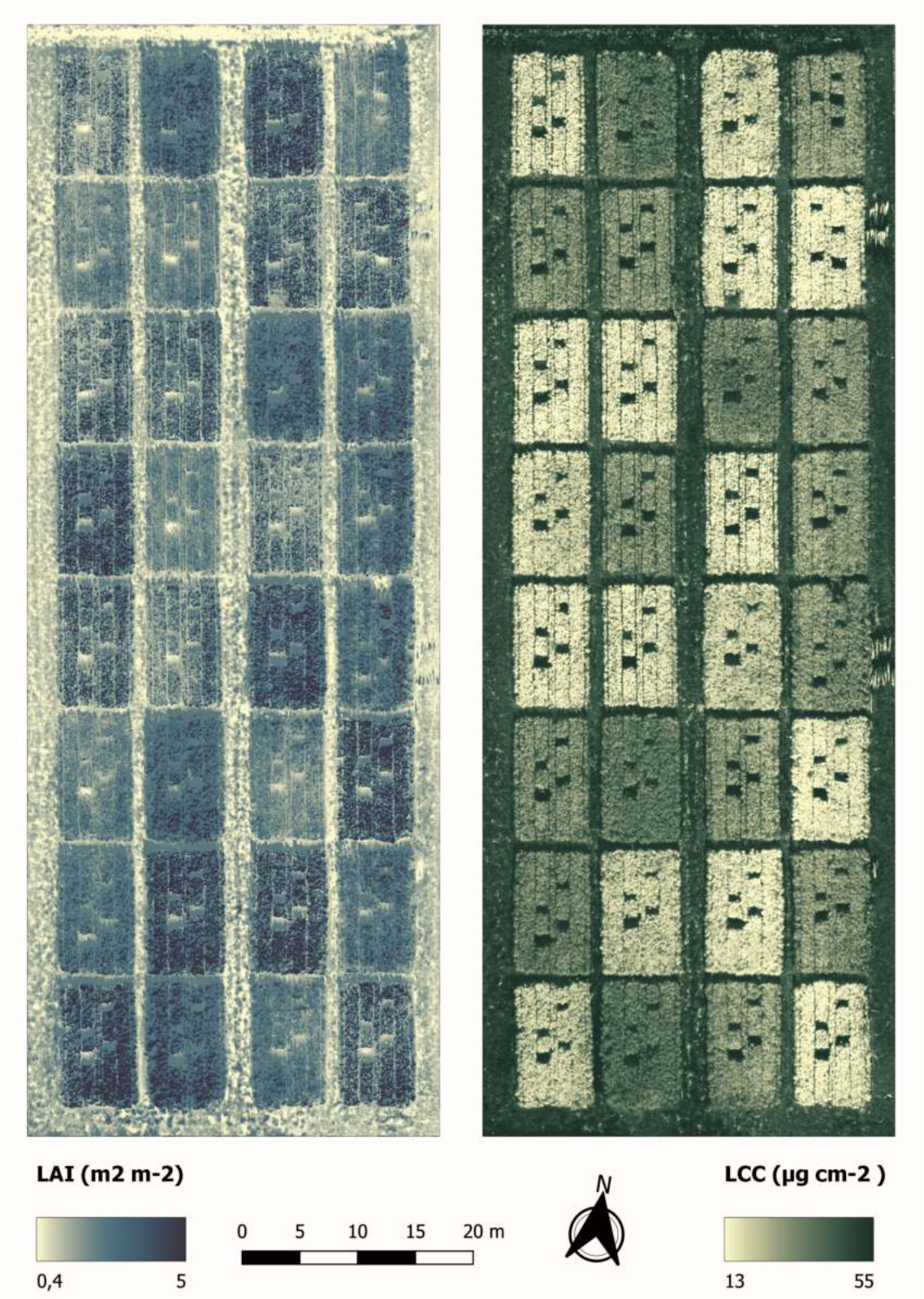


Figure 2.9 Maps of LAI and LCC estimated on 105 DAS during the growing season 2020 by UAV multispectral images using the best inversion of the PROSAIL model (RF for LAI and GPR for LCC).

Generalized additive model (GAM) was applied to the time series of LAI and LCC estimated by the inversion of the PROSAIL model, with 'Futura 75' as a reference for estimating significant differences among the hemp cultivars during the growing season (Figure 2.10). The estimated difference of LAI between 'Futura 75' and 'Fibror 79' was significant from 91 DAS to 108 DAS, with higher values of LAI for 'Fibror 79' than for 'Futura 75'; and from 129 DAS until the end of the growing season, but this time with higher values of LAI for 'Futura 75' than for 'Fibror 79' (Figure 2.10). Higher values of LAI were observed in 'Futura 75' than 'Fibror 79' during the early phases of the growing season (until 72 DAS), but no significant differences were observed. The estimated difference of LCC was significant throughout the whole growing season, with higher values of LCC for 'Futura 75' than for 'Fibror 79' (Figure 2.10). The estimated difference of LCC showed an increase from the start of the growing season up to 75 DAS (estimated difference of 13.5 µg cm^{-2}) and remained constant afterward until the end of the growing season.

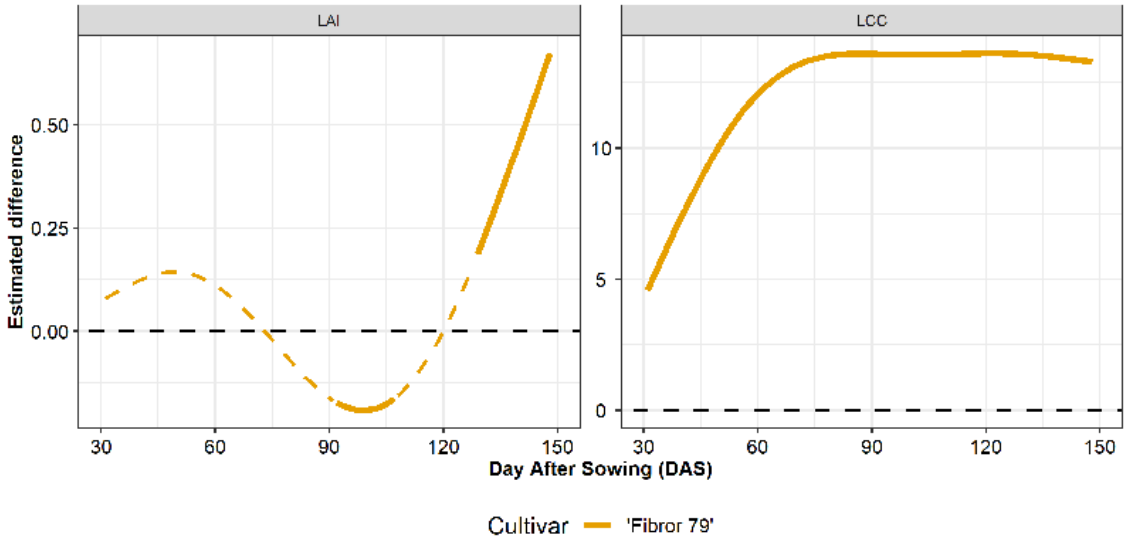


Figure 2.10 Traits' dynamics of the two hemp cultivars according to the difference in estimated LAI and LCC with reference the green cultivar 'Futura 75' (dashed black line). The estimation of the differences time series was carried out by using a GAM. Solid and dashed coloured line denotes respectively significant ($P<0.05$) and not significant differences of the corresponding 'Fibror 79' compared to reference 'Futura 75'.

2.3.4 Effect of nitrogen fertilisation on LAI and LCC dynamics

GAM was also used to analyse the effect of nitrogen fertilization on the dynamics of LAI and LCC of the two hemp cultivars. The reference used for estimating significant differences among nitrogen levels for each cultivar was the lowest fertilisation dose, 0 kg N ha^{-1} (Figure 2.11). The analysis showed that the effect of nitrogen dose was significant for both cultivars and trait. The largest

estimated differences were observed with the highest nitrogen level and decreased proportionally to the nitrogen dose. The estimated difference of LAI was highest during early phases of the growing season and progressively decreased until the final harvest (Figure 2.11). LCC showed the highest estimated difference at the end of the vegetative growth from 60 DAS to 90 DAS approximately. LAI of 'Futura 75' showed a significant estimated difference throughout the whole growing season with 100 kg N ha^{-1} nitrogen level and from the start of the growing season until 118 DAS and 103 DAS with 50 kg N ha^{-1} and 25 kg N ha^{-1} nitrogen levels, respectively. 'Futura 75' LCC showed a significant estimated difference from the start of the growing season until 144 DAS, 145 DAS and 82 DAS with 100 kg N ha^{-1}, 50 kg N ha^{-1} and 25 kg N ha^{-1} nitrogen levels, respectively (Figure 2.11). LAI of 'Fibror 79' showed a significant estimated difference from the start of the growing season until 140 DAS, 99 DAS and 70 DAS with 100 kg N ha^{-1}, 50 kg N ha^{-1} and 25 kg N ha^{-1} nitrogen levels, respectively. 'Fibror 79' LCC showed a significant estimated difference throughout the whole growing season with 100 kg N ha^{-1} nitrogen level, from the 37 DAS and 40 DAS until 102 DAS and 141 DAS with 50 kg N ha^{-1} and 25 kg N ha^{-1} nitrogen levels, respectively (Figure 2.11).

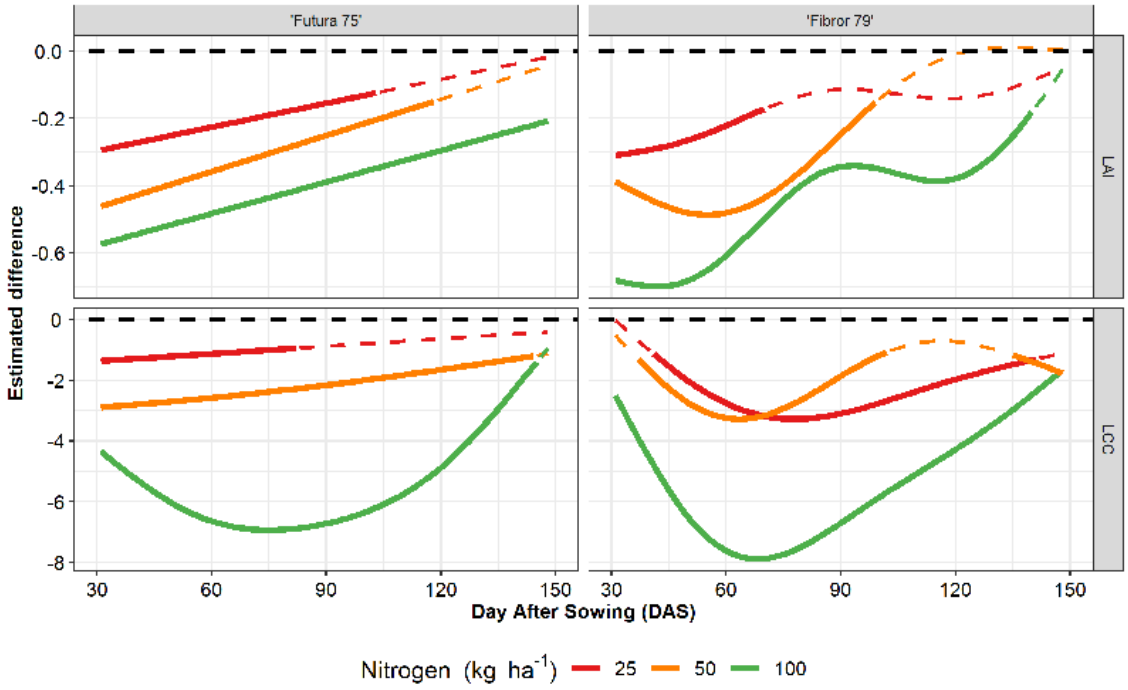


Figure 2.11 Dynamics of the estimated difference of LAI and LCC of the two cultivars with reference the nitrogen fertilisation level with 0 kg N ha^{-1} (dashed black line). The estimation of the differences time series was carried out by using a GAM. Solid and dashed coloured line denotes respectively significant ($P<0.05$) and not significant differences of the corresponding nitrogen fertilisation levels compared to reference nitrogen level with 0 kg N ha^{-1}.

2.4 Discussion

2.4.1 Evaluation of the inversion methods accuracy for the estimation of LAI and LCC

This study focused on UAV-based remote sensing estimation of hemp traits to phenotype two contrasting cultivars and to support the application of innovative precision farming.

Multispectral data acquired from unmanned aerial vehicle (UAV) in two growing seasons (2020 and 2021) were used to estimate leaf area index (LAI) and leaf chlorophyll content (LCC) using inversion methods of the PROSAIL model. The UAV multispectral camera used in this study (MicaSense RedEdge-Mx), which includes five spectral bands, enabled a reliable estimation of LAI and LCC, achieving results comparable to those obtained in previous studies, conducted with both multispectral (Sun *et al.*, 2021; Wan, Zhang, *et al.*, 2021; Zhu *et al.*, 2019) and hyperspectral UAV cameras (Duan *et al.*, 2014; Wang *et al.*, 2021). Among the different inversion methods present in literature, this study compared two look-up table (LUT) methods based on the $RMSE_r$ cost function (LUT-I and LUT-II), and four hybrid regression methods based on machine learning techniques (RF, GPR, ANN and EM) to identify the most accurate LAI and LCC estimation method. The results obtained with the LUT methods showed that the LUT-I method (single best solution) achieved a better accuracy than the LUT-II method (mean of 100 best solutions). Opposite results were reported, for heterogenous grassland, by Darvishzadeh *et al.* (2008) and for wheat by Sehgal *et al.* (2016). Regarding hybrid methods, the best accuracies were achieved with random forest (RF) for LAI and gaussian process regression (GPR) for LCC, which was also the most accurate method for estimating LCC in a study conducted on multi-crop by Verrelst *et al.* (2012).The worst accuracies were achieved by ensemble method (EM) for both LAI and LCC, which could be the consequence of the high correlation among the values predicted by the individual algorithms (RF, GPR and ANN), as already reported by Kamir *et al.* (2020).

In this research, estimation of LAI and LCC was more accurate using hybrid methods than LUT ones, which is in agreement with what was reported by Fei *et al.* (2012) and Zhang, Yang, *et al.* (2020), but not with findings of Sehgal *et al.* (2016) and Vohland *et al.* (2010) who found that the LUT inversion method was more accurate than hybrid methods. Ali *et al.* (2021), instead, did not find differences between the methods. However, accuracy is not the only criteria to consider when selecting the inversion method of the PROSAIL model. For example, studies have focused on the inversion run time, showing that hybrid methods are faster in performing the inversion compared to LUT methods (Ali *et al.*, 2021; Verrelst *et al.*, 2019; Verrelst, Camps-Valls, *et al.*, 2015).

In this study, LAI estimation was more accurate than the estimation of LCC. This confirms the difficulty in estimating LCC already found in previous studies (Baret & Jacquemoud, 1994; Darvishzadeh *et al.*, 2008; Sehgal *et al.*, 2016), which is the consequence of the poor signal propagation from leaf to canopy scale, as demonstrated by Asner (1998). The lowest accuracy of LCC estimation was observed under 20 µg cm^{-2}; this may have been caused by hemp bloom at the end of the season, which can affect reflectance data as found for rice cobs by Wan, Zhang, *et al.* (2021). On the other hand, analysing the accuracy of the LAI estimation, no relevant differences in NRMSE were observed in the two seasons (2020 and 2021) and in the three LAI intervals (< 2 m^2 m^{-2}, 2-4 m^2 m^{-2} and > 4 m^2 m^{-2}) for both cultivars. This proves that the inversion of the PROSAIL model can estimate LAI effectively, even on cultivars that have great LCC differences throughout the season.

2.4.2 UAV remote sensing and GAM for phenotyping the dynamics of LAI and LCC

High-throughput phenotyping (HTP) obtained by the combination of multiple UAV flights, estimation models of crop traits and generalized additive model (GAM) analysis can characterize the dynamics of relevant crop traits (Impollonia *et al.*, 2022). In particular, this study focused on the application of UAV multispectral images for characterising LAI and LCC of hemp. The time series of the traits estimated (with the best method for each trait) were used to characterise two hemp cultivars (a green and a yellow one), cultivated at different nitrogen fertilisation levels, using generalized additive model (GAM). The use of GAM to compare LAI dynamics of two hemp cultivars showed that LAI of 'Futura 75' tended to be higher (but not significantly) than 'Fibror 79' in the early phases of the growing season. At the end of the vegetative phase LAI was highest in 'Fibror 79' and finally it became highest in 'Futura 75' during the seed maturing phase (Figure 2.10). The highest LAI of 'Futura 75' at the start of the growing season indicates that its canopy develops more rapidly than that of ‘Fibror 79’. This is in accordance with the results obtained by Blandinières (2022) on the same trial, as they showed that 'Fibror 79' was slower than 'Futura 75' to reach canopy closure. After 90 days after sowing (DAS), however, the LAI of 'Fibror 79' increases faster than in 'Futura 75' and become even significantly higher for 18 days (between 91 DAS and 108 DAS). This can be explained by the fact that 'Fibror 79' is a cultivar that flowers slightly later than 'Futura 75', hence, the senescence of 'Futura 75' starts sooner than for 'Fibror 79', explaining the observed dynamic of LAI differences between both cultivars. These results are in accordance with Herppich *et al.* (2020) who reported that the LAI peak of ‘Ivory’ (the yellow cv.), a cultivar that flowers earlier than ‘Santhica 27’ (the green cv.), was reached sooner than the LAI peak of ‘Santhica 27’. They also found that for the green cultivar ‘Santhica 27’ LAI remained the highest for the rest of the growing season. After

flowering, the same dynamics was observed in this study where the LAI of the green cultivar 'Futura 75' was significantly higher than that of 'Fibror 79' during the seed filling phase. This could be due to the higher nitrogen content in 'Futura 75' leaves than in ‘Fibror 79’ ones. This could be due to the higher nitrogen content in 'Futura 75' leaves than in ‘Fibror 79’ ones, in line with the statement of Thouminot (2015) who stated that the yellow strain of hemp (i.e., Fibror 79) was due to a reduced capacity of nitrogen assimilation. Indeed, the dynamic of LCC observed in this study, linked to leaf nitrogen content (Wang *et al.*, 2014), showed higher values for ‘Futura 75’ than for ‘Fibror 79’ throughout the whole growing season. The LCC estimated differences of the two hemp cultivars showed an increase from the start of the growing season until the 75 DAS and then remained constant until the end of the growing season (Figure 2.10). In order to characterize the dynamics of the crop traits, the nitrogen fertilisation effects were included in the GAM analysis. Nitrogen fertilisation had a great and significant effect on the LAI and LCC dynamics (Figure 2.11). As for LAI, the higher difference, between no nitrogen fertilisation (reference) and the other nitrogen fertilisation levels, were observed in the early phases of the growing season for both cultivars. These results are in accordance with those reported by Seleiman *et al.* (2013), who found that the nitrogen fertilisation treatments had a significant effect on LAI of hemp only at the start of the growing season (44 DAS). The higher difference of LAI between the nitrogen levels could be due to a greater nitrogen accumulation at the start of the growing season, as reported by Ivonyi *et al.* (1997). They found, for hemp, that the most intense phase of nitrogen accumulation occurred between 30 and 60 DAS, as 79% of the total amount of nitrogen had effectively been accumulated after 60 DAS, in accordance with Seleiman *et al.* (2013). The intense nitrogen uptake during the early phases of the growing season could be explained by the general increase of differences of LCC until 60-80 DAS where was observed a peak of LCC differences (Figure 2.11). The increase of nitrogen fertilisation led to increases of nitrogen uptake and accumulation by the crop, with a subsequent significant increase of LCC, as reported by Yang *et al.* (2021). This relation was also observed in this study, as LCC dynamics had higher values at increasing levels of nitrogen fertilisation. After flowering, the LCC estimated differences decreased with the start of the senescence stage and the start of the chlorophyll’s degradation.

2.5 Conclusion

This study demonstrated that hemp traits can be estimated with good accuracy by the inversion of the PROSAIL model using multispectral images acquired by unmanned aerial vehicle (UAV). The leaf area index (LAI) can be estimated better than the leaf chlorophyll content (LCC). Generally, the hybrid methods performed better than look-up table (LUT) methods, both for LAI and LCC, and the

best accuracies were achieved by random forest (RF) for the LAI and gaussian process regression (GPR) for the LCC. The high throughput phenotyping (HTP) of the crops can be carried out by applying the generalized additive model (GAM) to the time series of traits estimated by the inversion of the PROSAIL model from multiple multispectral UAV flights. The GAM analysis showed differences in the LAI and LCC dynamics between two hemp cultivars, yellow and green, and between the nitrogen fertilisation levels that significantly affected the traits' dynamics. HTP based on UAV remote sensing proved a powerful tool to estimate crop traits and to improve our understanding of the traits' dynamics of contrasting cultivars throughout the whole growing season.

Chapter 3

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

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

Abstract

Miscanthus is a tall perennial C_4 rhizomatous grass with strong potential for sustainable biomass production and multiple co-benefits on marginal and contaminated lands, avoiding land competition with traditional food crops. Breeding programs in several countries are attempting to produce high-yielding *Miscanthus* hybrids better adapted to different climates and end-uses. Remote sensing with unmanned aerial vehicles (UAVs) equipped with multispectral sensors holds great potential for high-throughput phenotyping (HTP) of yield and quality traits. In this study, UAVs were flown fortnightly over ground phenotyped *Miscanthus* crops in Italy and the UK in 2020-21. Vegetation indices (VIs) derived from UAV multispectral images were used successfully to estimate yield traits (light interception, plant height, green leaf biomass and standing biomass) for diverse novel *Miscanthus* hybrids in Italy and in the UK using the random forest (RF) machine learning algorithm. RF algorithm and peak descriptor derived from time series of VIs were used to predict yield in order to improve the logistics biomass supply chain of *Miscanthus*. The Root Mean Square Errors (RMSE) of the RF models were 8.4 % for light interception, 42 cm for plant height, 1.3 Mg DM ha^{-1} for green leaf biomass, 5.8 Mg DM ha^{-1} for standing biomass estimations, and 2.3 Mg DM ha^{-1} for yield prediction. RF model showed a good capability to predict the yield months before the harvest both in Italy and in the UK. The PROSAIL model successfully linked VIs derived from different multispectral sensors on the UAVs used in Italy and UK. This 'interoperability' paves the way for linking VIs derived from different multispectral sensors carried by UAVs. The study demonstrates the potential for multispectral data from UAVs in HTP for genetic improvement and in the prediction of yield for providing important information needed to increase sustainable biomass production and to expand the bioeconomy.

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

Contents

3.1 Introduction

Miscanthus is a leading perennial biomass crop that can provide high yields on marginal lands avoiding land competition with traditional food crops (Amaducci *et al.*, 2017; Pancaldi & Trindade, 2020; Shepherd *et al.*, 2020). Yield is one of the most important traits of *Miscanthus* (Lewandowski *et al.*, 2000) and has been the primary focus of the research portfolio on *Miscanthus* in the last ten years (Clifton-Brown *et al.*, 2017; Clifton-Brown *et al.*, 2001; Jones *et al.*, 2016). Yield depends not only on the climatic and soil characteristics but also on crop age, harvest date and the genotype/hybrid type (Jones *et al.*, 2016). The rate of mature yield development varies with climate, but in general increases from 1-3 Mg ha^{-1} of dry matter (DM) in the first year to 6-8 Mg DM ha^{-1} in the second year and to 12 -30 Mg DM ha^{-1} in the third year and onwards in rainfed Northern and rainfed or irrigated Southern Europe (Clifton-Brown *et al*., 2007; Clifton-Brown *et al*., 2001; Jeżowski, 2008; Larsen *et al*., 2014). 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 (Clifton-Brown, Harfouche, *et al*., 2019; Lewandowski *et al*., 2016; van der Cruijsen *et al*., 2021). In the EU-BBI demo-project GRACE, we are evaluating novel highly upscalable seed-based *Miscanthus* hybrids (Clifton-Brown, Harfouche, *et al.*, 2019; Clifton-Brown, Schwarz, *et al.*, 2019; Hastings *et al.*, 2017; Pancaldi & Trindade, 2020; van der Cruijsen *et al.*, 2021) in seven European countries (Awty-Carroll *et al.*, 2022; Magenau *et al.*, 2022). Most yield prediction to date has relied on crop growth models driven by climate and soil data with crop specific parameters (de Wit & van Diepen, 2008; Keating *et al.*, 2003; MacKerron & Haverkort, 2004). For example, MISCANFOR is a crop growth model specifically developed to predict *Miscanthus x giganteus* yields in a wide range of environments (Hastings *et al.*, 2009). It has been widely used and validated at European (Lewandowski *et al.*, 2016) and national level (Zhang, Hastings, *et al.*, 2020) for *Miscanthus* and other perennial biomass crops (Henner *et al.*, 2020), but new parameterization data to predict yield production of the novel *Miscanthus* hybrids (Clifton-Brown, Harfouche, *et al.*, 2019) is required (Hastings *et al.*, 2009).

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 (Liu *et al*., 2019; Richter *et al*., 2016). Impollonia *et al.* (2022) 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) (Blancon *et al.*, 2019; Makanza *et al.*, 2018; Potgieter

et al., 2017), ii) calibrate crop growth models (Jongschaap, 2006; Prévot *et al.*, 2003) and iii) predict the yield of many crops for commercial purposes (Ferchichi *et al.*, 2022). Crop traits such as the plant height (Jin *et al.*, 2018), the fraction of absorbed photosynthetically active radiation (fAPAR) (Upreti *et al.*, 2019) and the aboveground biomass (Han *et al.*, 2019) 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) (Adam *et al.*, 2014; Verrelst *et al.*, 2019; Wang *et al.*, 2016). RF proved to be robust to outliers and noise, does not suffer from overfitting, and can manage a high training size (Belgiu & Drăguţ, 2016). The use of ML algorithms shows great potentials in crop yield prediction (Hunt *et al.*, 2019; Jeong *et al.*, 2016; Marques Ramos *et al.*, 2020; Senthilnath *et al.*, 2016). 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) (de Beurs & Henebry, 2005; Ji *et al.*, 2021; Meroni, d'Andrimont, *et al.*, 2021) such as the start of season (SOS), the peak of the growing season, the stay-green duration (onset of senescence), the end of the season (EOS), and the growing season length (de Beurs & Henebry, 2010). 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 (Montazeaud *et al*., 2016) and EVI2 (Liu *et al*., 2019) for grain yield and the peak of GNDVI for biomass yields of perennial grass (Hamada *et al*., 2021).
However, the VIs values are influenced by many factors, such as sensor characteristics, atmospheric conditions during acquisition, viewing angle, field of view, and sun elevation (Psomiadis *et al*., 2017). 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 (Théau *et al*., 2010). 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 (Psomiadis *et al.*, 2017). 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 (Emilien *et al.*, 2021; Hoque & Phinn, 2018). The multi-sensor interoperability is an important topic in remote sensing science (Brown *et al.*, 2006; Gallo *et al.*, 2005; Meroni *et al.*, 2013) when multi-location monitoring is conducted. There are two main approaches to evaluate and compare VIs derived from different sensors via linear regression (She *et al.*, 2015; Teillet *et al.*, 2006; Teillet & Ren, 2008): a direct approach where VIs are measured by sensors (Laliberte *et al.*, 2011) and an indirect approach where VIs data are simulated by radiative transfer models (van Leeuwen *et al.*, 2006). A major limitation of the direct regression approach is that is not transferable because it is

site-specific (Hoque & Phinn, 2018). On the contrary, the indirect approach permits to retrieve VIs data from radiative transfer models such as the PROSAIL model (Baret *et al.*, 1992; Jacquemoud *et al.*, 2009). PROSAIL has been used to assess the performances of different satellite sensors for multiple VIs (Verrelst, Camps-Valls, *et al.*, 2015) but a similar application is currently lacking in UAV science. Regarding the yield predictions, it 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. Furthermore, to date, few studies have focused on the estimation and prediction of perennial biomass crops phenotypic traits using remote sensing technologies from satellite or UAV (Hamada *et al.*, 2021; Li *et al.*, 2020; Wang *et al.*, 2019) and only two on *Miscanthus* (Impollonia *et al.*, 2022; Richter *et al.*, 2016).

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 canopy reflectance and to link VIs of two different common multispectral sensors, ii) random forest (RF) algorithm to estimate crop traits and to predict yield using the VIs and peak of VIs time series and iii) generalized additive model (GAM) to derive peak from complete and partial time series of VIs.

3.2 Materials and methods

3.2.1 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) (Figure 3.1). These sites are two of the seven plot scale (PS) trials conducted within EU-BBI GRACE demo-project (Awty-Carroll *et al.*, 2022). 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 (Table 3.1) with n=4 replicates for a total of n= 56 plots. The trials were planted with eight novel intraspecific *M. sinensis x M. sinensis* hybrids *(M. sin x M.*

sin), five novel interspecific hybrids *M. sinensis x M. sacchariflorus* (*M. sin x M. sac*) and *M. x giganteus* as control genotype (for more details see Awty-Carroll *et al.* (2022)).

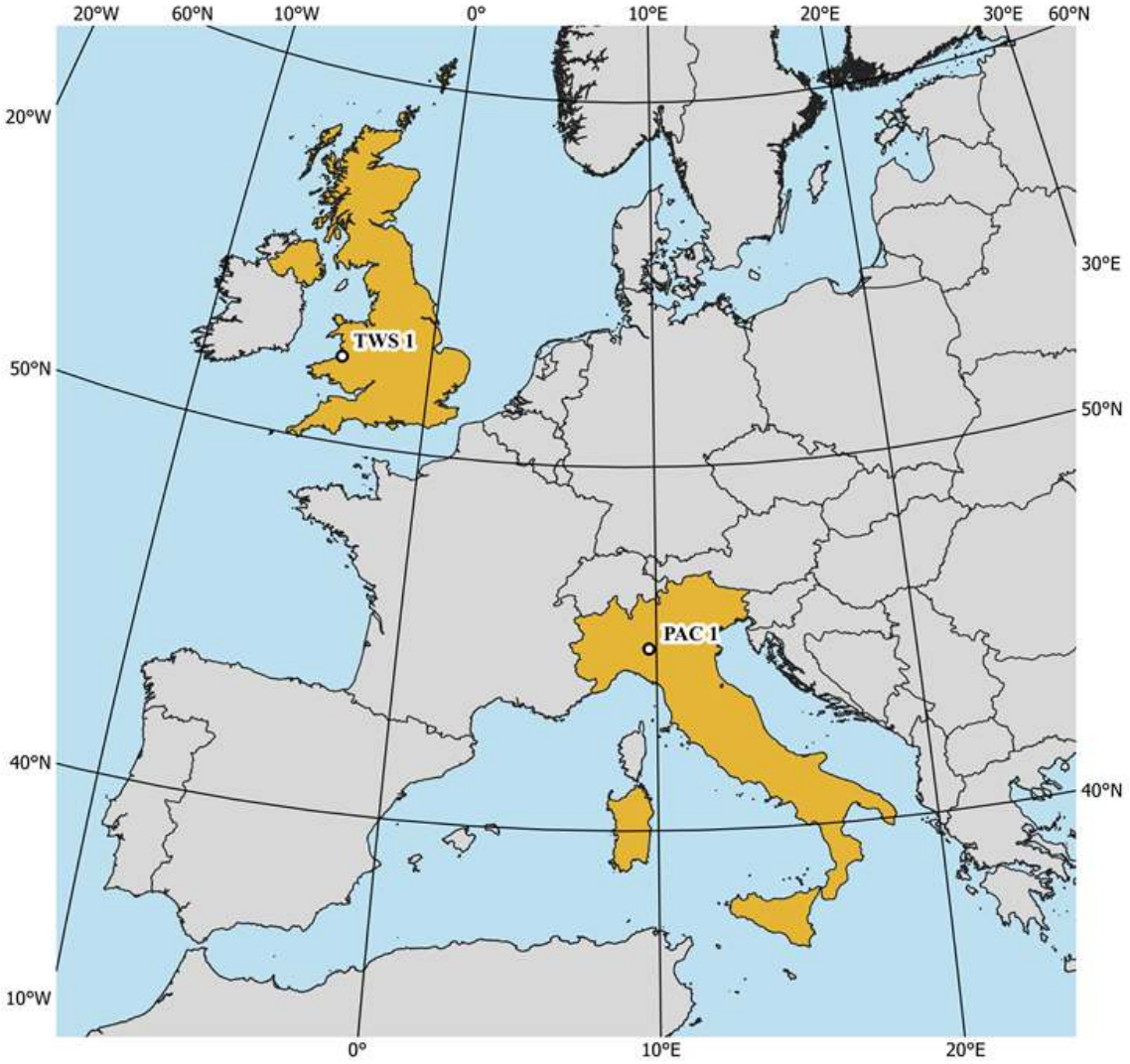


Figure 3.1 Map of field experiments: PAC 1 is located in Piacenza (North-West Italy) and TWS 1 is located in Aberystwyth (Mid-West Wales).

Table 3.1 Characteristics of the 14 *Miscanthus* hybrids considered in this study.

Material	Hybrid code	Genotype	Planting density
Seed-based plugs	GRC 1-8	*M.sinensis x M. sinensis*	3 plants/m^2
Rhizomes *	GRC 9	*M. x giganteus*	1.5 plants/m^2
Seed-based plugs	GRC 10 - 14 (except GRC 12)	*M.sinensis x M. sacchariflorus*	1.5 plants/m^2
Rhizomes *	GRC 15	*M.sinensis x M. sacchariflorus*	1.5 plants/m^2

* Hybrids commercially available

3.2.2 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 x M. sin*), GRC 14 (seeded) and GRC 15 (clonal) *(M. sin x M. sac*) and GRC 9 (standard clonal *M. x giganteus*). The measurements of this study were carried during 3rd growing season. Five contiguous plants along a central row in each plot were used for monitoring along the growing season (see Awty-Carroll *et al.* (2022) for a detailed description). 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^{-1}) and standing biomass (Mg DM ha^{-1}) 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.* (2022). 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^{-1}) 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 80ºC were used to calculate the mass and moisture contents of each fraction and were scaled to Mg DM ha^{-1}. The crop phenological stages were estimated using the thermal time following the method proposed on *M. x giganteus* by Tejera *et al.* (2021). Thermal time was measured in growing degree days (GDD, °C Day, Equation 3.1) as:

$$GDD = \left[\frac{T_{max} + T_{min}}{2}\right] - T_b \qquad \text{Equation 3.1}$$

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 (Farrell *et al*., 2006). 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 in spring was measured for all the fourteen hybrids using a quadrat area of 6.6 m^2 (10 plants per plot planted at 1.5 plants m^{-2} (*M. x giganteus* and *M. sin x M. sac*) or 20 plants per plots at 3 plants m^{-2} (*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 subsample of approximately five stems per plot was used to determine the moisture content and calculate the yield in Mg DM ha^{-1}. Plants were harvested on February 2nd 2021 (Days of year (DOY): 33) at PAC 1 and on March 8th 2021 (DOY: 67) at TWS 1. A full description of the harvest protocol as well as annual yield data from winter harvest used in this study is provided in Awty-Carroll *et al.* (2022).

3.2.3 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 3.2). Table 3.2. details the specifications of the UAVs and the multispectral cameras used at two sites. All 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 at 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, Switzerland). The 15 vegetation indices (VIs) were calculated as shown in Table 3.3 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, USA) and georeferenced with QGIS software (QGIS Development Team, 2021).

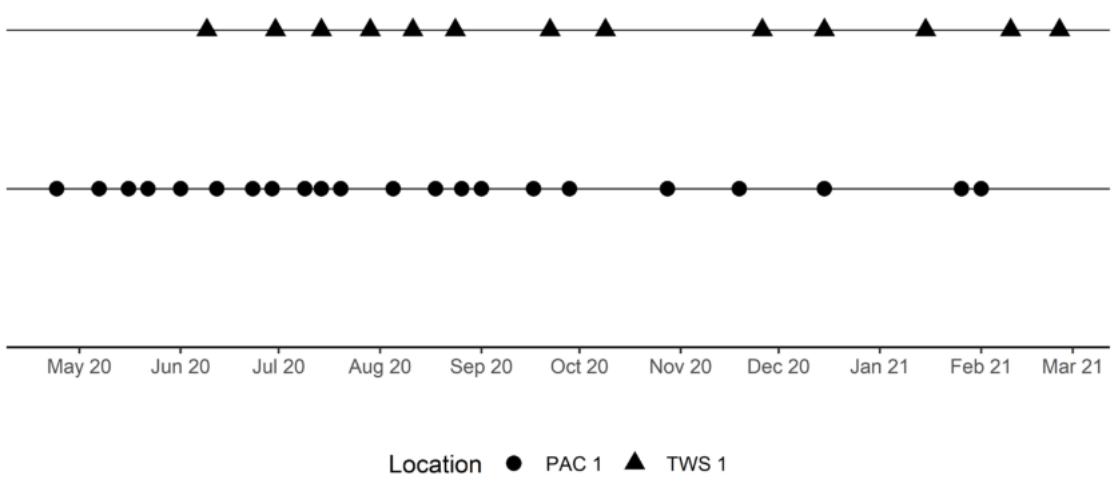


Figure 3.2 Seasonal calendar of UAV flights performed in the two locations.

Table 3.2 Unmanned aerial vehicles (UAVs) and multispectral cameras used to perform flights in the two locations.

Location	UAV	Camera	Spectral bands
PAC 1	DJI M210 RTK	MicaSense RedEdge-Mx	blue (475 nm centre, 32 nm FWHM), green (560 centre, 27 nm FWHM), red (668 nm centre, 14 nm FWHM), red edge (717 nm centre, 12 nm FWHM), near-infrared (840 nm centre, 57 nm FWHM)
TWS 1	DJI M210	SlantRange 4P	blue (470 nm centre, 100 nm FWHM), green (550 centre, 100 nm FWHM), red (650 nm centre, 40 nm FWHM), red edge (710 nm centre, 20 nm FWHM), near-infrared (850 nm centre, 100 nm FWHM)

Table 3.3 List of the vegetation indices evaluated in random forest models for crop traits estimation and for yield prediction of *Miscanthus*.

VIs	Equation	Reference
Datt1	$\frac{\text{NIR} - \text{RedEdge}}{\text{NIR} + \text{Red}}$	Datt, 1999
EVI2	$2.4\frac{\text{NIR} - \text{Red}}{1 + \text{NIR} + \text{Red}}$	Miura *et al.*, 2008
GNDVI	$\frac{\text{NIR} - \text{Green}}{\text{NIR} + \text{Green}}$	Gitelson *et al.*, 1996
GOSAVI	$\frac{\text{NIR} - \text{Green}}{\text{NIR} + \text{Green} + 0.16}$	Sripada *et al.*, 2006
greenWDRVI	$\frac{0.1\text{NIR} - \text{Green}}{0.1\text{NIR} + \text{Green}} + \frac{1 - 0.1}{1 + 0.1}$	Gitelson, 2004
MSAVI	$\frac{2\text{NIR} + 1 - \sqrt{(2\text{NIR} + 1)^2 - 8(\text{NIR} - \text{Red})}}{2}$	Qi *et al.*, 1994
MTVI1	1.2(1.2(NIR - Green) - 2.5(Red - Green))	Haboudane *et al.*, 2004
MTVI2	$1.5\frac{1.2(\text{NIR} - \text{Green}) - 2.5(\text{Red} - \text{Green})}{\sqrt{(2\text{NIR} + 1)^2 - 6\text{NIR} - 5\sqrt{\text{Red}} - 0.5}}$	Haboudane *et al.*, 2004
NDRE	$\frac{\text{NIR} - \text{RedEdge}}{\text{NIR} + \text{RedEdge}}$	Gitelson & Merzlyak, 1994
NDVI	$\frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red}}$	Rouse *et al.*, 1973
OSAVI	$(1 + 0.16)\frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red} + 0.16}$	Rondeaux *et al.*, 1996
OSAVI2	$(1 + 0.16)\frac{\text{NIR} - \text{RedEdge}}{\text{NIR} + \text{RedEdge} + 0.16}$	Wu *et al.*, 2008
rededgeWDRVI	$\frac{0.1\text{NIR} - \text{RedEdge}}{0.1\text{NIR} + \text{RedEdge}} + \frac{1 - 0.1}{1 + 0.1}$	Gitelson, 2004
SAVI	$(1 + 0.5)\frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red} + 0.5}$	Huete, 1988
WDRVI	$\frac{0.1\text{NIR} - \text{Red}}{0.1\text{NIR} + \text{Red}} + \frac{1 - 0.1}{1 + 0.1}$	Gitelson, 2004

3.2.4 Using the PROSAIL model to link VIs from different multispectral sensors

The PROSAIL model was used to simulate crop spectral signatures to link VIs (Table 3.3) calculated from the two different multispectral sensors used in this study (Table 3.2). 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 (Rusinowski *et al*., 2019; Urrego *et al.*, 2021). The *hsdar* R package (Lehnert et *al.*, 2019) 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 3.4.

Table 3.4 Ranges of input parameters for the PROSAIL model for the generation of the LUT.

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

Regarding the parameter combinations used for LUT generation, all parameter combinations were considered in the analysis as novel *Miscanthus* hybrids were evaluated (no information on literature) and the crop monitoring was performed throughout the whole growing season (vegetative growth and senescence). The spectral reflectance simulated were resampled based on UAV sensors characteristics (Table 3.2) and the 15 VIs used in this study were calculated. For each VI, a linear regression was realised to link 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 VIs that need a linking procedure and those that do not. The final database of VIs was built by linking SlantRange VIs data toward MicaSense one.

3.2.5 Time series of VIs and peak derivation

The 15 VIs calculated from UAV multispectral images of the two sites and linked using the linear regression derived by the PROSAIL model were smoothed using a generalized additive model (GAM) to generate daily time series of the VIs. The GAM is a non-parametric regression model which allows non-linear fitting of the variables. GAM models were fitted in R package *mgcv* (Wood, 2017). GAM fitting of VIs allows to remove the outliers and regularize the time series (Antonucci *et al.*, 2021; Impollonia *et al.*, 2022; Nolè *et al.*, 2018). The time series of VIs for each plot were fitted against the modified days of the year (DOY). The modified DOY was used to overcome the problem of having non-sequential DOY throughout the whole growing season, as this last overlaps two different years. Modified DOY were calculated as conventional DOY for the first year of the growing season (2020) and as DOY + 365 starting from the 1st of January for the year 2021.

Daily time series of the VIs were used to estimate the crop traits and to derive peak descriptor for each VI to predict yield. The peak descriptor is defined as the maximum value recorded in the time series of a VI. Two types of VIs time series were evaluated in the peak derivation: i) complete VIs time series, where the GAM models were fitted using all the VIs data acquired throughout the whole growing season, and ii) partial VIs time series, where the GAM models were fitted with fixed VIs data acquired in the first seven UAV flights (175 DOY in PAC 1 and 266 DOY in TWS 1) and adding the subsequent single UAV flight until 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 operability for yield prediction during *Miscanthus* growing season.

3.2.6 Machine learning modelling and variable importance

The random forest (RF) algorithm was used to estimate the crop traits (section 3.2.2) and to predict the yield of *Miscanthus* hybrids. The RF models were created using the *caret* R package (Kuhn, 2008). Three steps in RF modelling were adopted: firstly, RF models were created using the 15 VIs of the 4 *Miscanthus* hybrids (see section 3.2.2) and phenological stages (encoded as 0 and 1 for the stages of vegetative growth and senescence, respectively) for traits estimation, and the peak derived from complete time series of 15 VIs of the 14 *Miscanthus* hybrids for yield prediction; secondly, the importance of variables in 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 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 (Han *et al.*, 2019): 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, Woźniak *et al.*, 2021) using the DALEX package (Biecek, 2018). To show the uncertainty of importance estimation, the variable importance was calculated for 10 permutations (Woźniak *et al.*, 2021). The RF models' performances were evaluated calculating the root mean square error (RMSE, Equation 3.2) and the normalized root mean square error (NRMSE, Equation 3.3) as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_i - y_i)^2}{n}} \qquad \text{Equation 3.2}$$

$$NRMSE\ (\%) = \frac{\sqrt{\frac{\sum_{i=1}^{n}(x_i - y_i)^2}{n}}}{\bar{y}} 100 \qquad \text{Equation 3.3}$$

where n is the sample number, x_i and y_i are the estimated and measured value of each trait, and $\bar{y}$ is the mean of the measured value. 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.

3.3 Results

3.3.1 PROSAIL model for linking VIs from different multispectral sensors

The canopy reflectance simulated by PROSAIL for the two multispectral sensors were used to link, via linear regression and on two locations, the vegetation indices (VIs). The outputs of the linear regressions for each VI calculated from PROSAIL model between the two sensors (*y* - the MicaSense VIs and *x* - the SlantRange VIs) are reported in Figure 3.3. EVI2, MSAVI and SAVI were the three VIs with the slope values closest to 1 and intercept values closest to 0 (Figure 3.3). 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 with SlantRange sensor, underestimating the VI compared to MicaSense (Figure 3.3). 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.

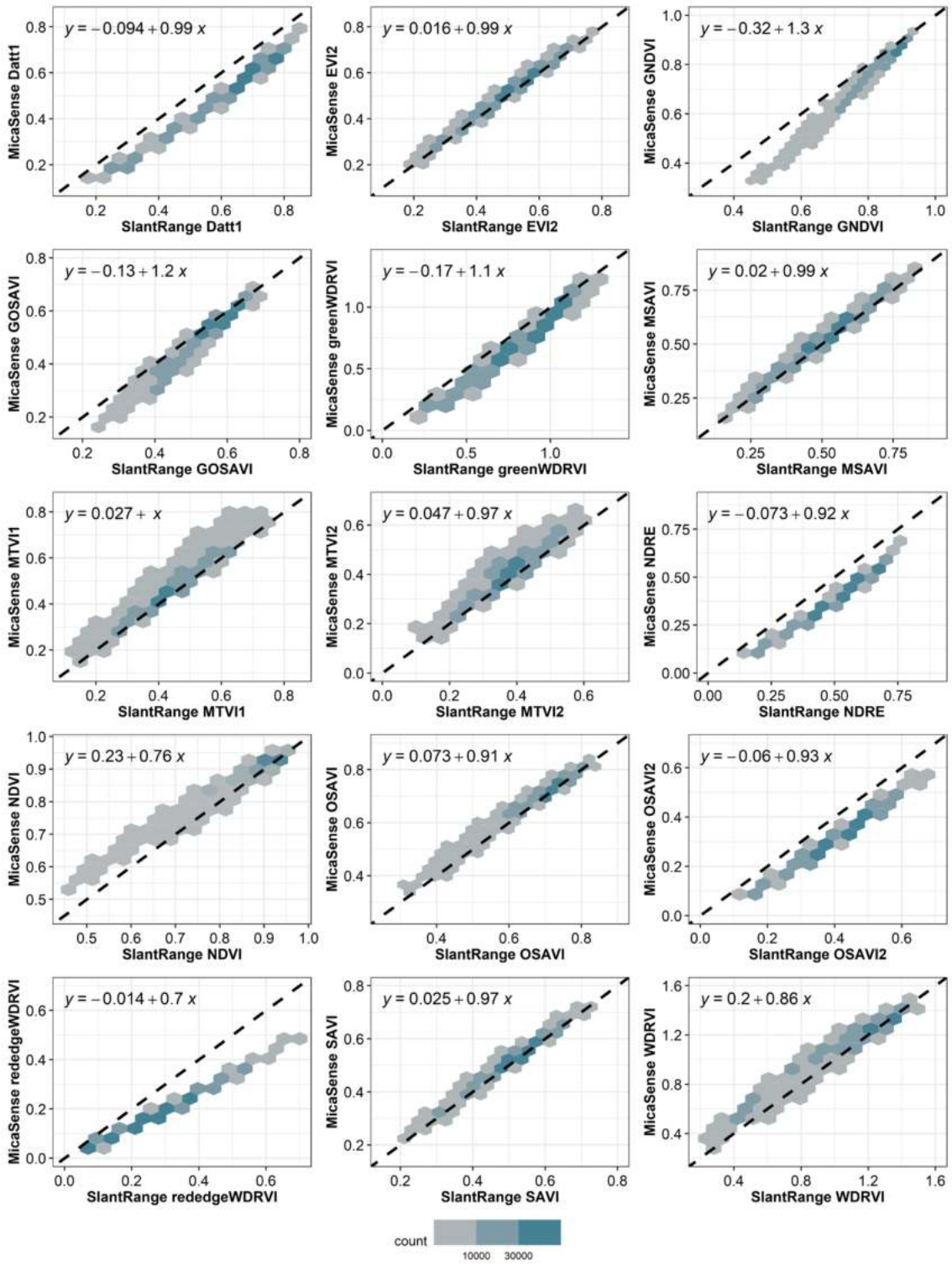


Figure 3.3 Linear regression between VIs MicaSense and SlantRange derived from the PROSAIL model. The 1:1 relationship is represented by a dashed line.

3.3.2 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 3.4. 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 (Biecek, 2018; Woźniak *et al*., 2021). Phenological stage ("Stage") was the most important variable for estimating plant height, green leaf biomass and standing biomass (Figure 3.4). For the crop traits estimation, 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, the most important VIs (with a median of RMSE loss greater than 0) of peak descriptor were greenWDRVI, NDVI, WDRVI, GNDVI and MTVI2 (Figure 3.4).

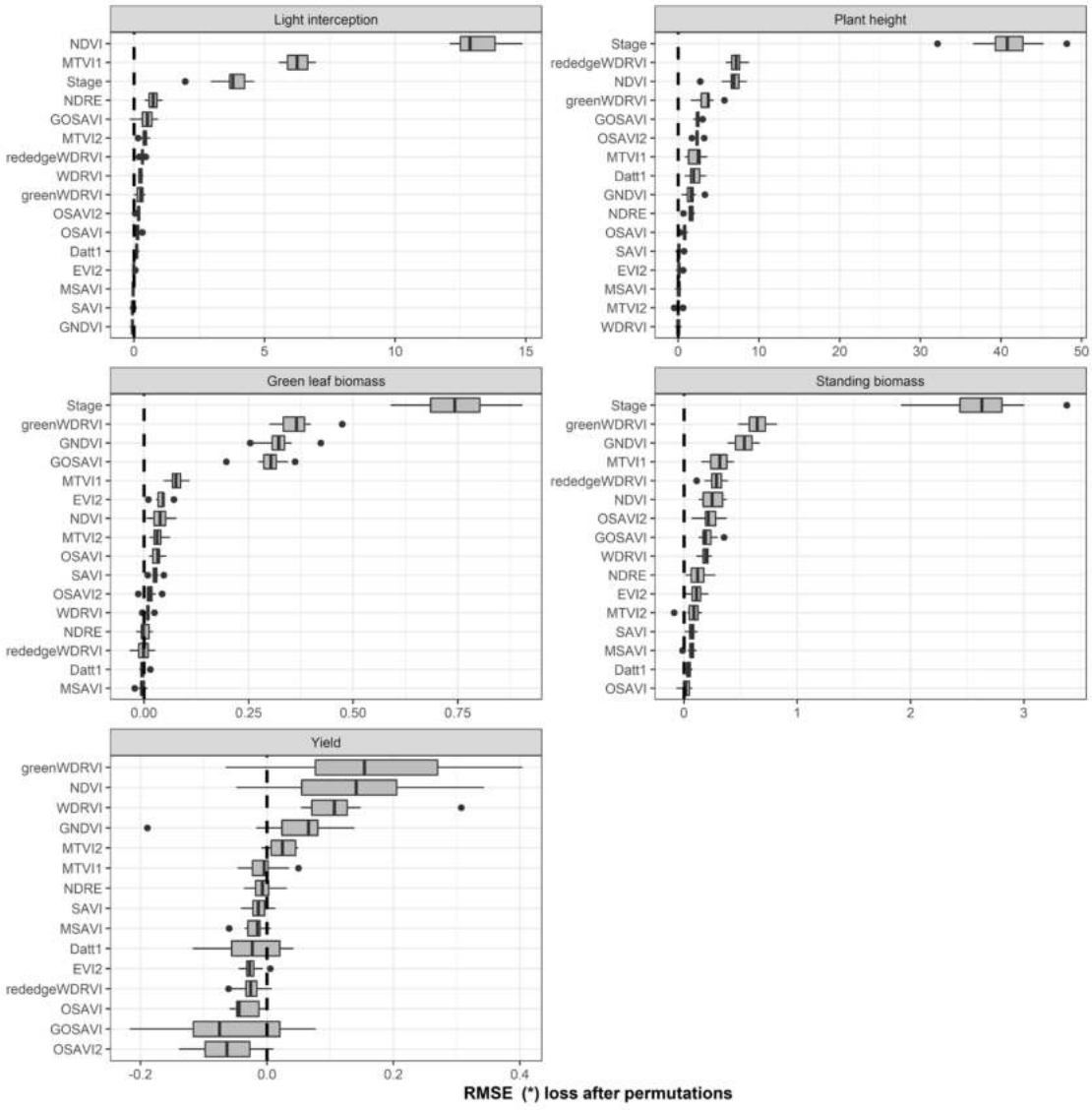


Figure 3.4 Importance of the RF models variables for crop trait estimations and for yield prediction, expressed as drop-out loss of model performance (RMSE) for each variable related to the drop-out loss of the full model (dotted line). *The RMSE values are in (%), (cm), (Mg DM ha^{-1}), (Mg DM ha^{-1}) and (Mg DM ha^{-1}) respectively for the light interception, plant height, green leaf biomass, standing biomass and yield.

3.3.3 Machine learning model for crop traits estimation

Crop traits distribution of contrasting *Miscanthus* hybrids measured at two locations is shown in Figure 3.5. 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 3.5). 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^{-1} to 14.5 Mg DM ha^{-1} and 5 Mg DM ha^{-1} in PAC 1 and from 0.05 Mg DM ha^{-1} to 6.3 Mg DM ha^{-1} and 1.5 Mg DM ha^{-1} in TWS 1. For standing biomass, the range was from 0.5 Mg DM ha^{-1} to 46.4 Mg DM ha^{-1} in PAC 1 and from 0.5 Mg DM ha^{-1} to 21.1 Mg DM ha^{-1} in TWS 1, and the mean was 8 Mg DM ha^{-1} in TWS 1 and 18.9 Mg DM ha^{-1} in PAC 1.

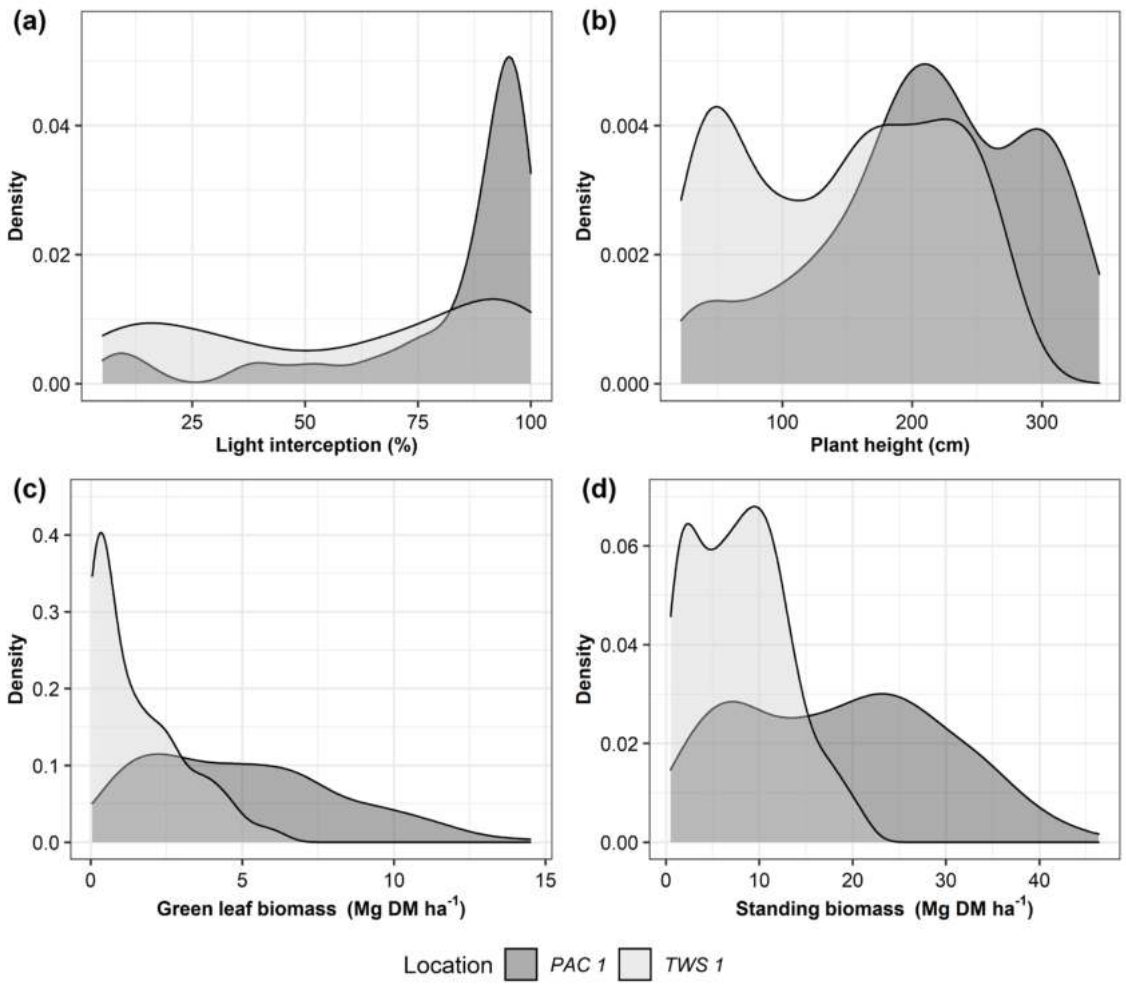


Figure 3.5 Frequency distribution of *Miscanthus* traits at the two locations PAC 1 and TWS 1: (a) light interception (%), (b) plant height (cm), (c) green leaf biomass (Mg DM ha^{-1}) and (d) standing biomass (Mg DM ha^{-1}).

Overall, the RF model estimated phenotypic traits with good model performances (Figure 3.6 and Figure 3.7). Among the crop traits, light interception was estimated with the greatest accuracy (NRMSE of 10.9 %, Figure 3.6). 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 3.6). For these last parameters, the RF model showed good accuracy from low to intermediate values, while above values of 5 Mg DM ha^{-1} of green leaf biomass and 20 Mg DM ha^{-1} of standing biomass the model performances dropped. The NRMSE performance metrics for each location and hybrid are reported in Figure 3.7. Generally, no relevant differences were observed between the two locations (Figure 3.7). 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.

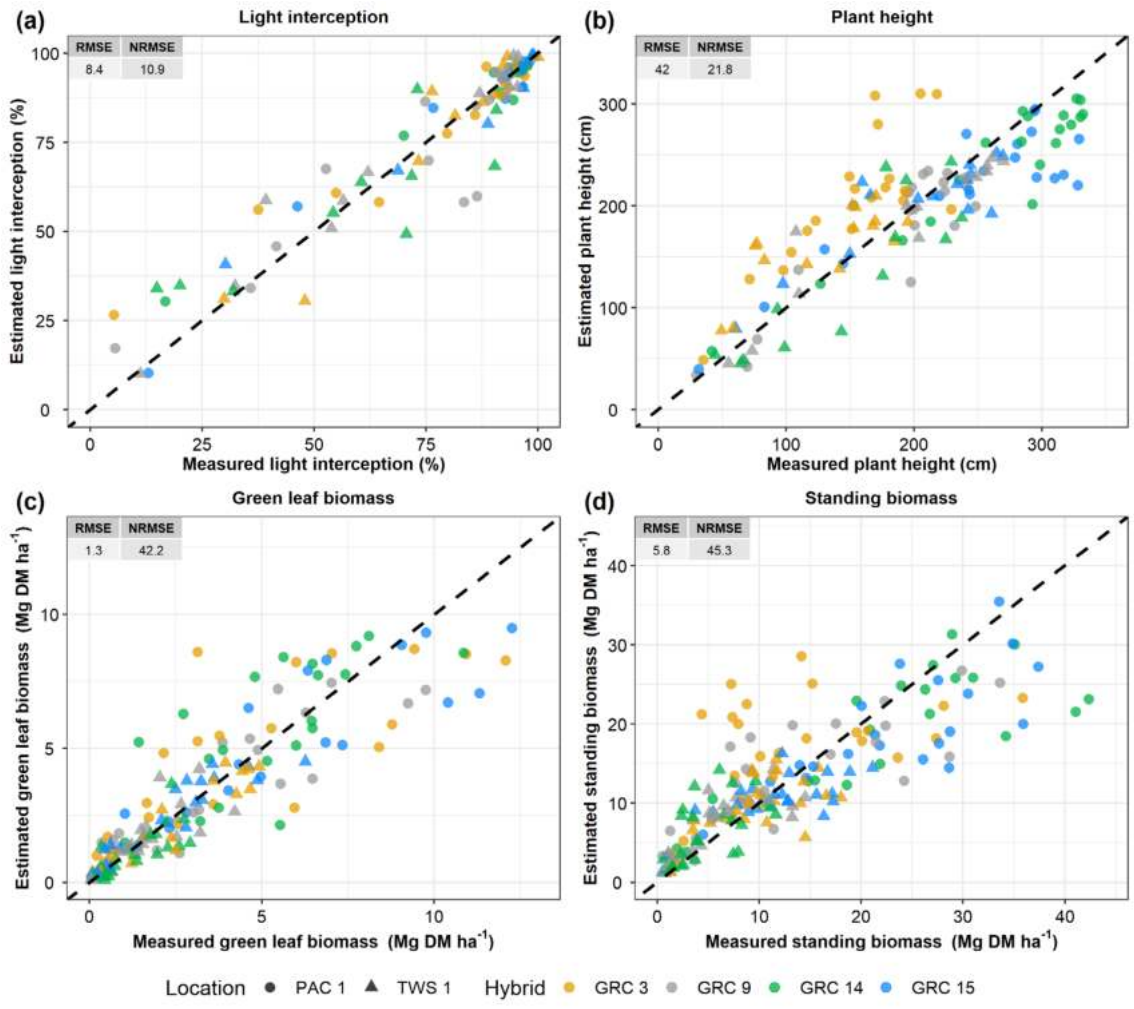


Figure 3.6 Estimated vs measured crop traits on the ground of four *Miscanthus* hybrids growth at the PAC 1 and TWS 1: (a) light interception (%), (b) plant height (cm), (c) green leaf biomass (Mg DM ha^{-1}) and (d) standing biomass (Mg DM ha^{-1}).

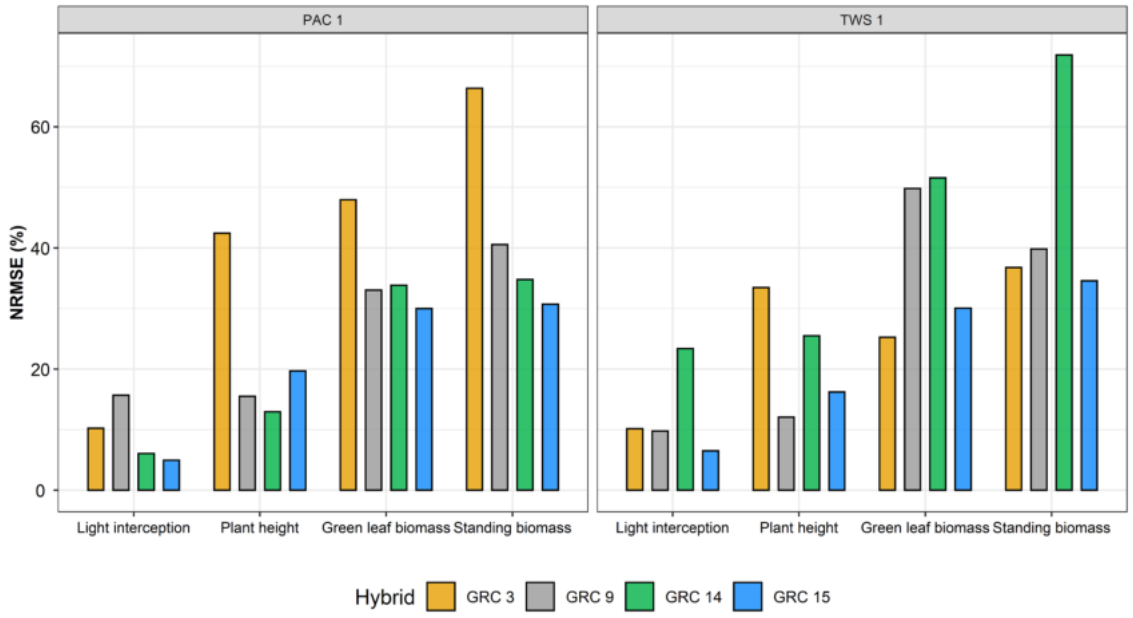


Figure 3.7 NRMSE values of the RF models obtained for each crop trait assessed in four hybrids grown at the two locations PAC 1 and TWS 1.

3.3.4 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 3.8. On average, the yield of *Miscanthus* was higher in PAC 1 than TWS 1. In PAC 1, the highest yield was recorded by *M. sin x M. sac* (mean of 18.3 Mg DM ha^{-1}). *M. sin x M. sin* productivity averaged 11.3 Mg DM ha^{-1}, while *M. giganteus* was the less productive (mean of 9.6 Mg DM ha^{-1}). In TWS 1, the highest yield was recorded by *M. sin x M. sin* (mean of 9.4 Mg DM ha^{-1}). *M. sin x M. sac* productivity averaged 8.2 Mg DM ha^{-1} while *M. giganteus* was the less productive (mean of 6.6 Mg DM ha^{-1}). The RF model, trained and tested with the yield values reported in Figure 3.8, enabled a reliable estimation of *Miscanthus* yield for all hybrids using the peak derived from complete time series of VIs (Figure 3.9). The RF model obtained a RMSE value of 2.3 Mg DM ha^{-1} and NRMSE value of 19.7 % (Figure 3.9a). 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 3.9b). The days of the year (DOY) of the peak of the VIs (greenWDRVI, NDVI, WDRVI, GNDVI and MTVI2) are reported in Figure 3.10. On average, the VIs reached the peak earlier in PAC 1 (172 DOY – June 20th) than in TWS 1 (DOY 232– August 19th). 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 - August 15th), 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 3.10).

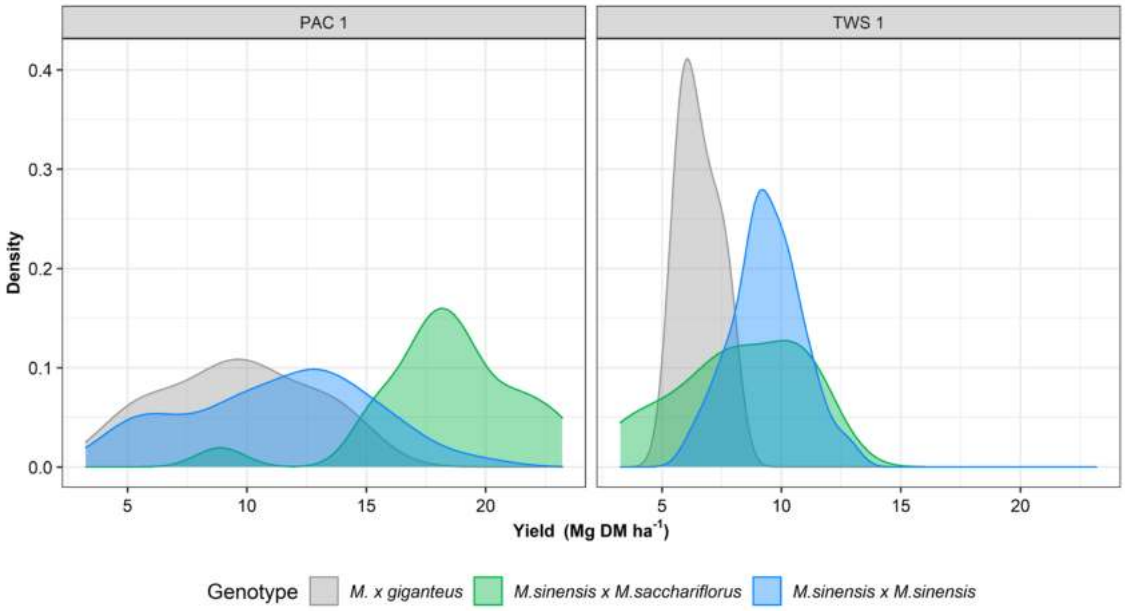


Figure 3.8 Frequency distribution of yield (Mg DM ha^{-1}) for the different *Miscanthus* hybrids at the two locations PAC 1 and TWS 1.

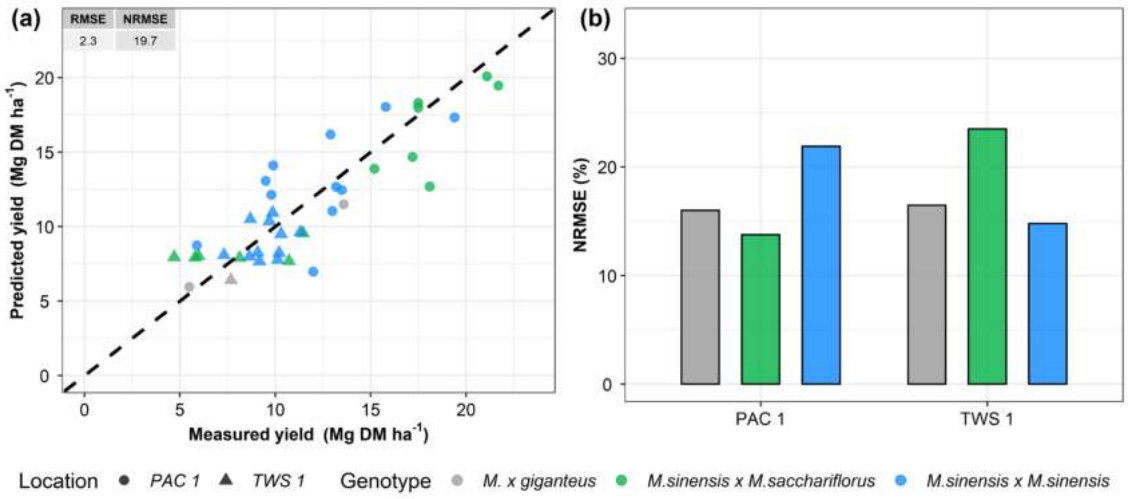


Figure 3.9 (a) Predicted vs measured yield and (b) NRMSE of the RF model for yield prediction of the different *Miscanthus* hybrids at two locations PAC 1 and TWS 1.

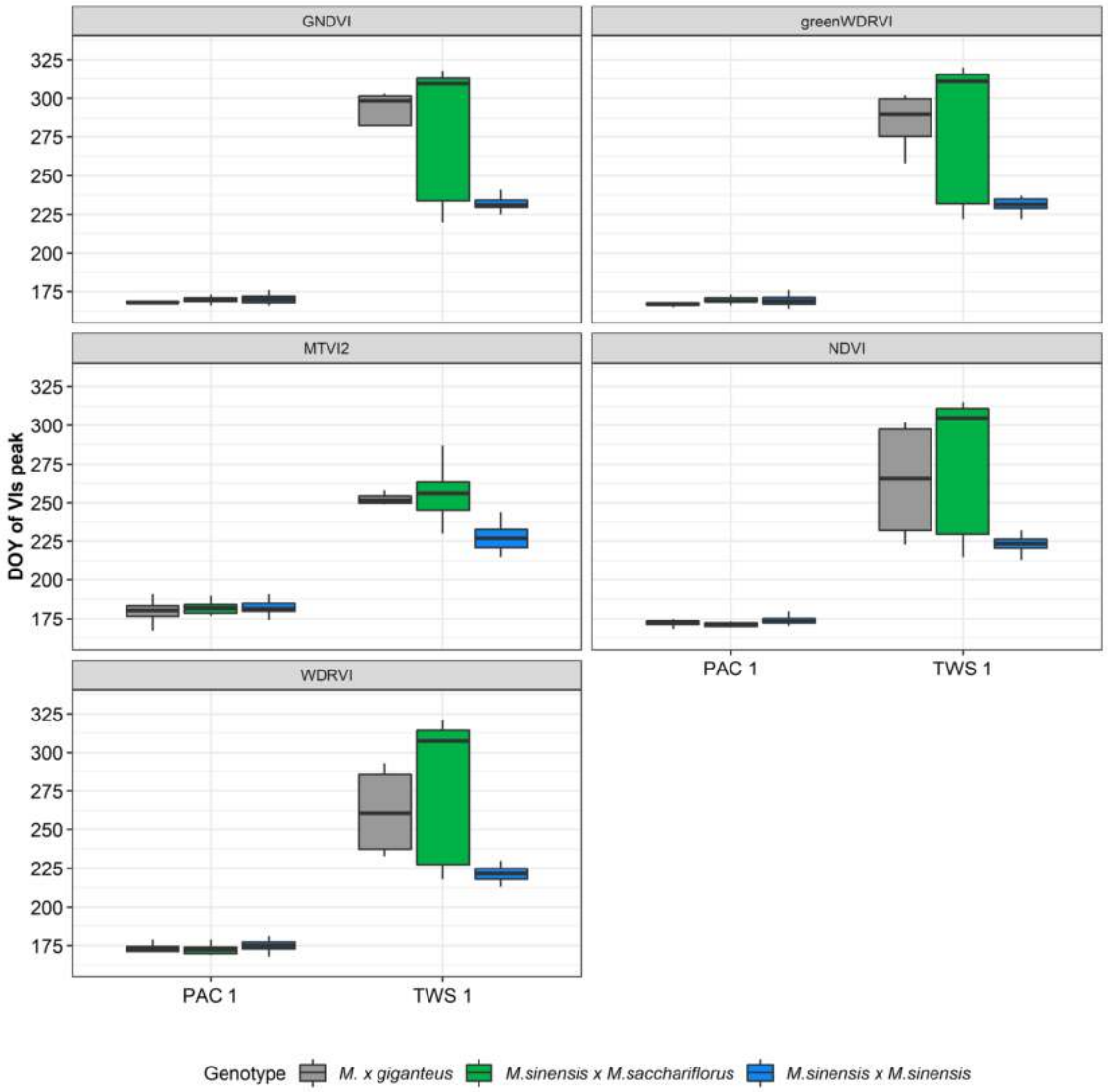


Figure 3.10 Boxplot of the day of the year (DOY) of the peak derived from complete time series of the greenWDRVI, NDVI, WDRVI, GNDVI and MTVI2, for the different *Miscanthus* hybrids at two locations PAC 1 and TWS 1.

3.3.5 Time series of VIs and yield prediction analysis

The complete time series of the 5 VIs selected (yield in Figure 3.4) for yield prediction are reported in Figure 3.11a for PAC 1 and Figure 3.12a 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 3.11a), following the same order of the mean yield measured in the field (Figure 3.8). In TWS 1, similar time series of all VIs were recorded for the *M. sin x M. sac* and *M. giganteus* (Figure 3.12a). In particular, the peaks of the *M. sin x M. sac* and *M. giganteus* were reached later than *M. sin x M. sin* as shown in Figure 3.10. The variation throughout the season of the peak derived by fitting the VIs via generalized additive model (GAM) are displayed in Figure 3.11b for PAC 1 and Figure 3.12b for TWS 1. In PAC 1, the differences 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 302 modified days of the year (DOY) of the UAV flight (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 3.11b). 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 331 modified DOY of the UAV flight (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 3.12b) in TWS 1 before 331 modified DOY. The timeline of the performance of the RF model tested with the VIs peak from partial time series is reported in Figure 3.13. In PAC 1, the NRMSE decreased until 302 modified DOY 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 modified DOY and onward. In TWS 1, no relevant differences in NRMSE were observed between the UAV flights performed from 266 modified DOY and onward.

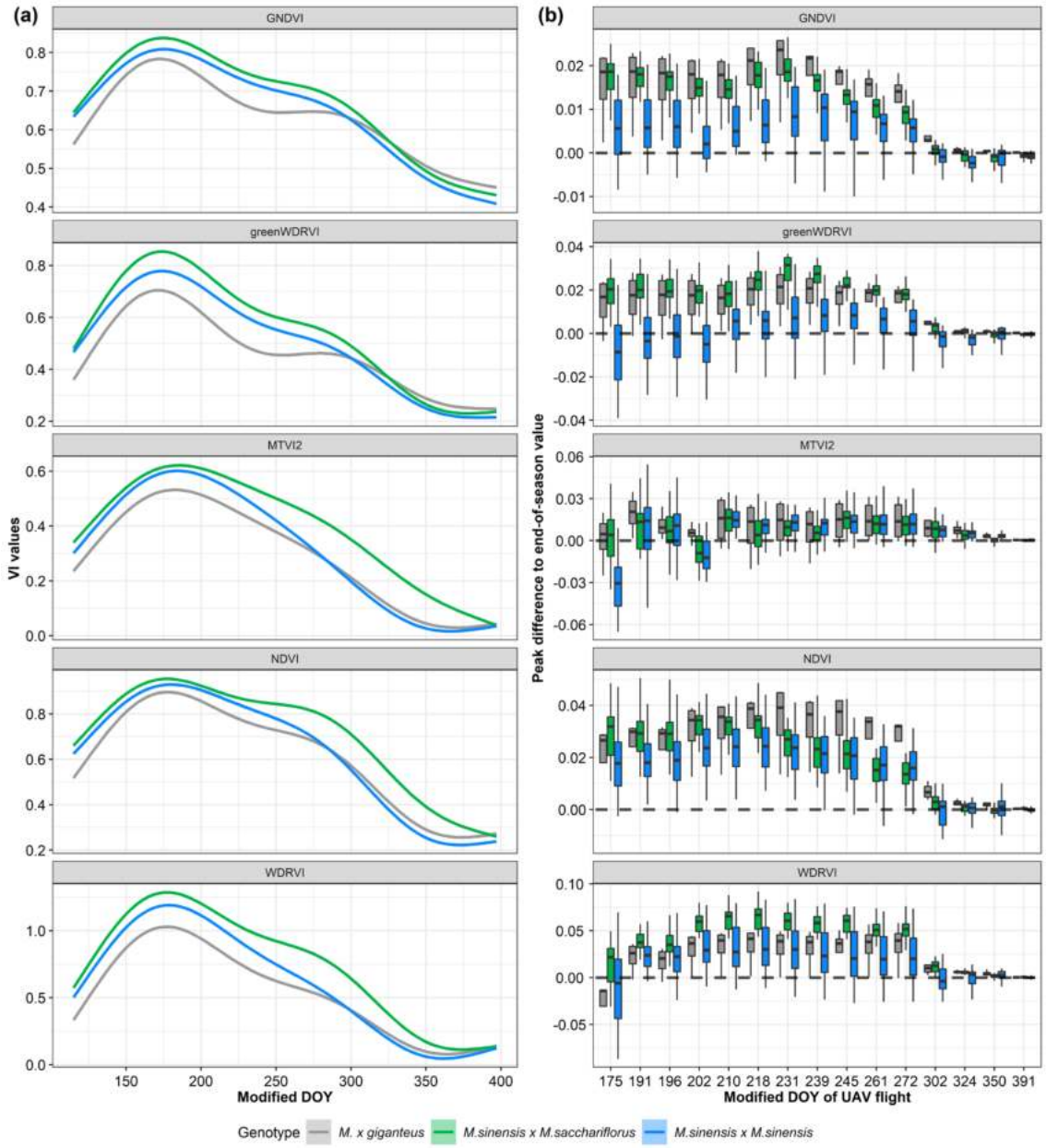


Figure 3.11 (a) Time series of the five VIs (GNDVI, greenWDRVI, MTVI2, NDVI, WDRVI) fitted via generalized additive model (GAM) throughout the whole growing season of *Miscanthus* in PAC 1. Modified day of the year (DOY) were calculated by adding 365 to the DOY of the corresponding year from January 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 modified DOY 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 modified DOY) and the peak derived from partial time series fitted until the modified DOY of the UAV flight.

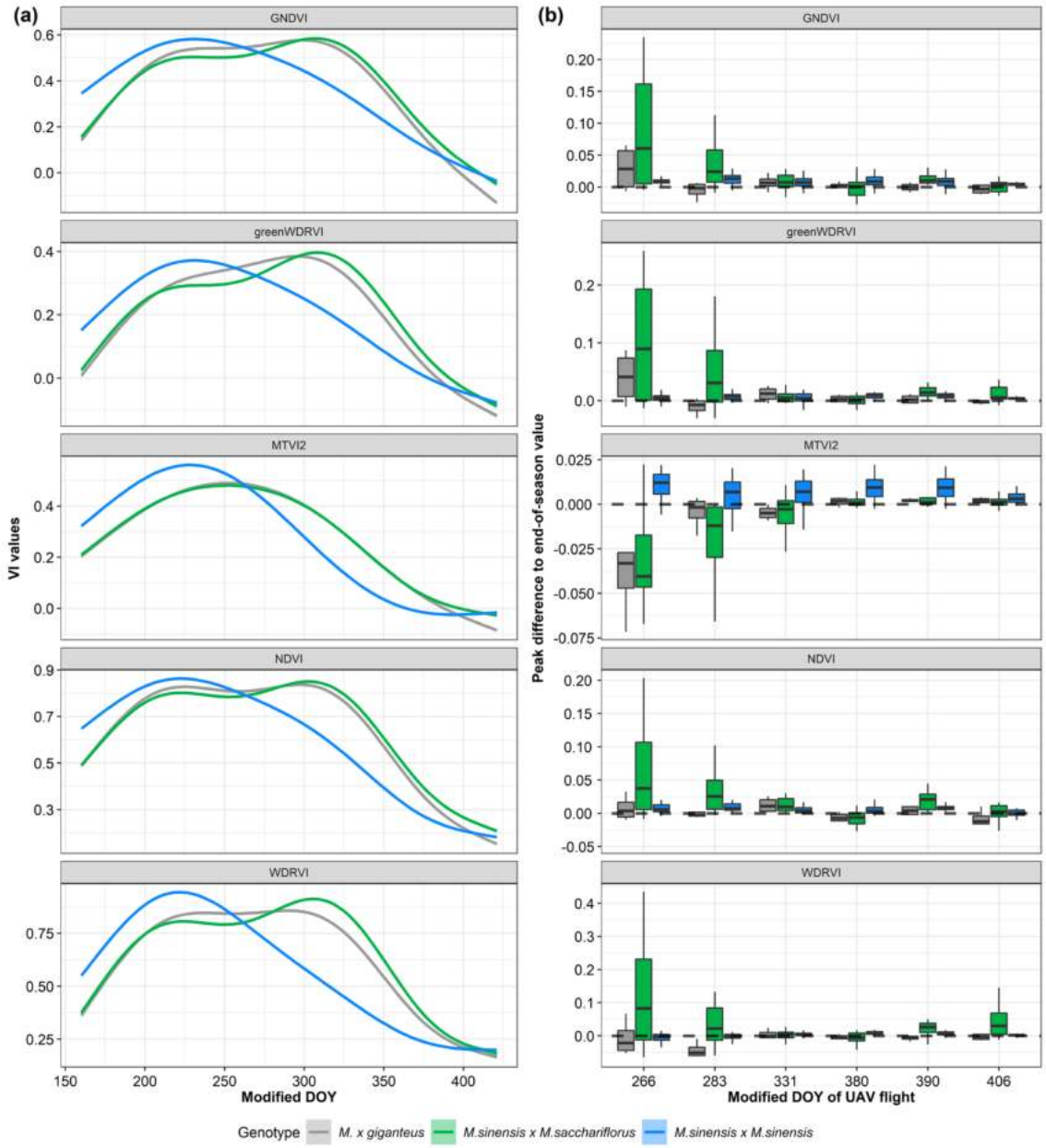


Figure 3.12 (a) Time series of the five VIs (GNDVI, greenWDRVI, MTVI2, NDVI, WDRVI) fitted via generalized additive model (GAM) throughout the whole growing season of *Miscanthus* in TWS 1. Modified day of the year (DOY) were calculated by adding 365 to the DOY of the corresponding year from January 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 modified DOY 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 modified DOY) and the peak derived from partial time series fitted until the modified DOY of the UAV flight.

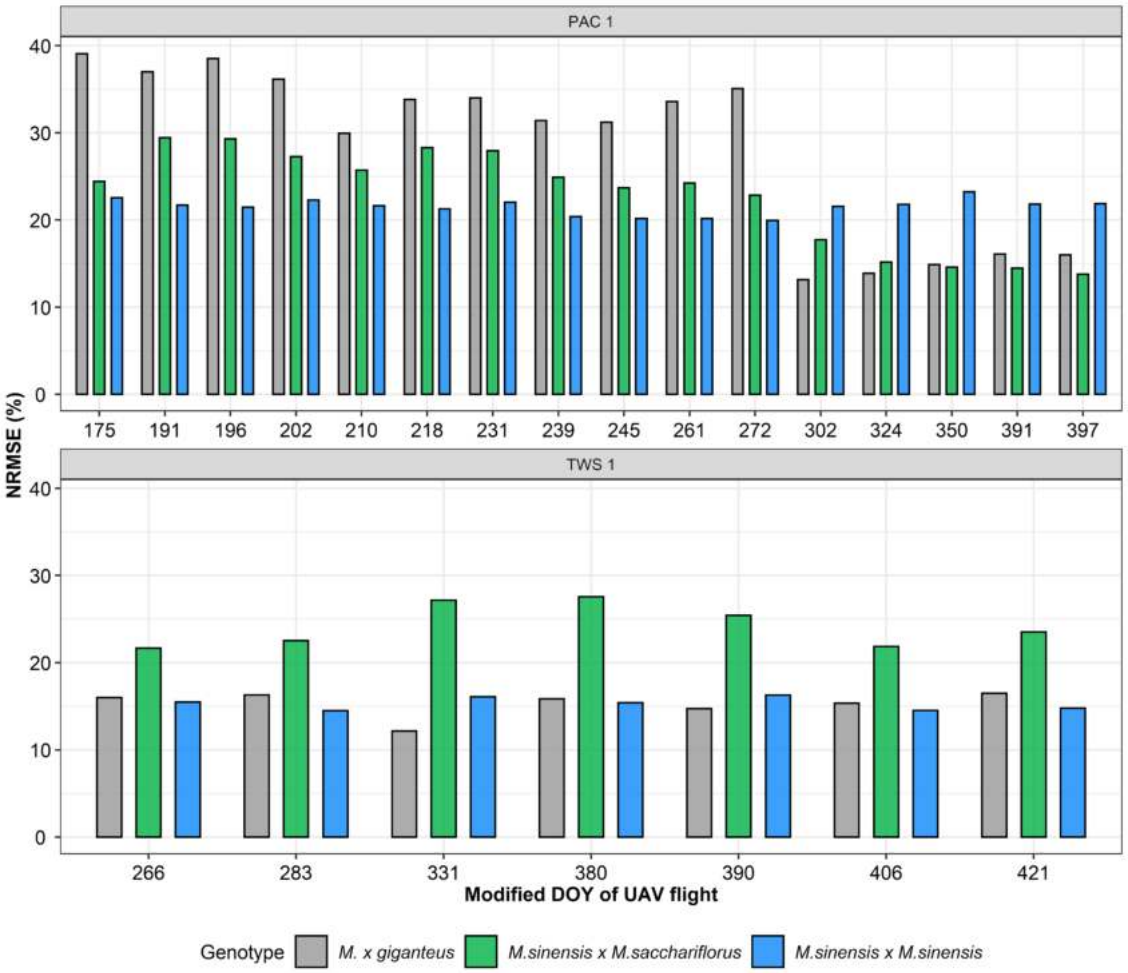


Figure 3.13 NRMSE values of the RF model for yield prediction, trained with the peak of the VIs derived from complete time series and tested with the peak of the VIs derived from partial time series fitted until the modified DOY of the UAV flight, at the two locations PAC 1 and TWS 1.

3.4 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).

3.4.1 The importance of linking VIs for multi-sensor interoperability

Linking VIs of multi-sensor is relevant for remote sensing crop monitoring (Meroni *et al.*, 2013), 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 link VIs calculated from two common multispectral cameras (MicaSense RedEdge-MX and SlantRange 4P). Ideally, the VIs from any sensor can be then linked against the VIs of a selected reference sensor. This linking approach is commonly adopted for VIs obtained from different satellites (van Leeuwen *et al.*, 2006) 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 correction for the same VIs were reported by Li *et al.* (2013) 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 linking 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 (Cui & Kerekes, 2018; Rengarajan & Schott, 2018).

This study highlighted the importance of linking outputs from different multispectral sensors to increase interoperability in remote sensing. Kim *et al.* (2010) and Villaescusa-Nadal *et al.* (2019) reported that the use of linear regression equations to link 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 linking 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 (Théau *et al*., 2010), other factors cause differences of VIs such as the atmospheric conditions during acquisition (Psomiadis *et al*., 2017). In the UAV

images acquisitions, the changing light and meteorological conditions during the flights can affect the quality of the spectral data (Guo *et al*., 2019). Therefore, this procedure is limited by other factors that cannot be considered by applying a simulated regression coefficient. However, using equations able to link 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.

3.4.2 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.* (2020) 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 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 3.6). This result is in agreement with Guillen-Climent *et al.* (2014) who found that the fraction of intercepted photosynthetically active radiation (fIPAR) was successfully estimated by a ML algorithm. Upreti *et al.* (2019) 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 3.6). A similar RMSE value (41 cm) was found by Han *et al.* (2019) for plant height estimation of maize using the crop surface model, and by Tao *et al.* (2020) (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^{-1} and 5.8 Mg DM ha^{-1} of RMSE, respectively (Figure 3.6). However, for these parameters, the model showed good accuracy from low to intermediate values, while above values of 5 Mg DM ha^{-1} of green leaf biomass and 20 Mg DM ha^{-1} of standing biomass the model performed dropped. The model could be affected by errors in the estimation in these intervals due to fewer data used to train the model (Shah *et al.*, 2019), in fact, most of them were collected in PAC 1 (Figure 3.5). The different levels of accuracy of the models in estimating light interception and plant height compared to green leaf biomass and standing biomass could be also related to the period in which 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 (Tillack *et al.*, 2014) while, as shown by Awty-Carroll *et al.* (2022) and Magenau *et al.* (2022), *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 3.4). 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 (Awty-Carroll *et al.*, 2022). *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^{-2}) and flowered earlier (end of August) compared to other genotypes that flowered in early autumn (Magenau *et al.*, 2022). This more "prostrate" canopy architecture (with a higher stem segment between inflorescences and bent leaves) introduced noise in the plant height estimation from UAVs (Volpato *et al.*, 2021) 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 3.7). In fact, the earlier a genotype with prostrate architecture flowers, the worse is the estimation of plant height and biomass from UAV.

3.4.3 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 RF model accurately predicted the yield with 2.3 Mg DM ha^{-1} of RMSE and 19.7 % NRMSE (Figure 3.9a). The peak of the five VIs used for yield prediction were selected by dropout loss of RMSE and the most important ones were the peak of the greenWDRVI, NDVI, WDRVI, GNDVI and MTVI2 (Figure 3.4). 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.* (2021), 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.* (2016) who found a high correlation between the peak of NDVI and the yield, and by Liu *et al.* (2019), 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 3.4) 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 3.4). Similar results for the GNDVI were found in switchgrass and other warm-season perennial grasses (Hamada *et al.,* 2021). In order to assess the capability of the model to predict the yield months before harvest, in this study, a timeline of the performance of the RF model was calculated using the peak derived from partial time series of VIs. Given the good accuracy achieved by the RF model (Figure 3.13), the results showed that yield can be predicted some months before harvest as already reported for switchgrass (Hamada *et al.*, 2021), both in Italy and in the UK, using multispectral images acquired from multiple UAV flights. . 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, implying to perform UAV multispectral image acquisition of *Miscanthus* up to end-October. 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.

3.5 Conclusion

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. High-throughput phenotyping and yield prediction 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 linking equations derived from the PROSAIL model proved to be a powerful tool to link VIs from multi-sensor with different spectral characteristics. Although this procedure is relevant for the upscale of models from experimental plots to field by linking 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.

Chapter 4

Moisture content estimation and senescence phenotyping of novel *Miscanthus* hybrids combining UAV based remote sensing and machine learning

This chapter is based on:

Impollonia, G.[1,2*], Croci, M.[1,2], Martani, E.[1], Ferrarini, A.[1], Kam, J.[3], Trindade, L. M.[4], Clifton-Brown, J.[5,6], Amaducci, S.[1,2] (2022). Moisture content estimation and senescence phenotyping of novel *Miscanthus* hybrids combining UAV based remote sensing and machine learning. GCB Bioenergy. https://doi.org/10.1111/gcbb.12930

[1]*Department of Sustainable Crop Production, Università Cattolica del Sacro Cuore, Piacenza 29122, Italy*

[2]*Remote Sensing and Spatial Analysis Research Center (CRAST), Università Cattolica del Sacro Cuore, Piacenza 29122, Italy*

[3]*Terravesta, Unit 4 Riverside Court, Skellingthorpe Road, Saxilby, Lincoln LN1 5AB*

[4]*Wageningen University & Research, Plant Breeding, 6700 AJ Wageningen, The Netherlands*

[5] *Aberystwyth University, Institute of Biological, Environmental and Rural Sciences, Aberystwyth, SY23 3EE, UK*

[6] *Justus-Liebig-Universität Gießen. iFZ, Heinrich-Buff-Ring 26, 35392 Gießen, Germany*

*Corresponding author

Moisture content estimation and senescence phenotyping of novel *Miscanthus* hybrids combining UAV based remote sensing and machine learning

Abstract

Miscanthus is a leading perennial biomass crop that can produce high yields on marginal lands. Moisture content is a highly relevant biomass quality trait with multiple impacts on efficiencies of harvest, transport and storage. The dynamics of moisture content during senescence and overwinter ripening are determined by genotype x environment interactions. In this chapter, unmanned aerial vehicle (UAV) based remote sensing was used for high-throughput phenotyping (HTP) of the moisture content dynamics during autumn and winter senescence of 14 contrasting hybrid types (progeny of *M. sinensis x M. sinensis* (*M. sin x M. sin*, 8 types) and *M. sinensis x M. sacchariflorus* (*M. sin x M. sac*, 6 types)). The time series of moisture content was estimated using machine learning (ML) models and a range of vegetation indices (VIs) derived from UAV based remote sensing. The most important VIs for moisture content estimation were selected by the recursive feature elimination (RFE) algorithm and were BNDVI, GDVI and PSRI. The ML model transferability was high only when the moisture content was above 30%. The best ML model accuracy was achieved by combining VIs and categorical variables (5.6% of RMSE). This model was used for phenotyping senescence dynamics and identifying the stay-green (SG) trait of *Miscanthus* hybrids using the generalized additive model (GAM). Combining ML and GAM modelling, applied to time series of moisture content values estimated from VIs derived from multiple UAV flights, proved to be a powerful tool for HTP.

Keywords: *Miscanthus*, moisture content, remote sensing, UAV, multispectral, machine learning, transferability, senescence, GAM, high-throughput phenotyping.

Contents

4.1 Introduction

Miscanthus is a promising perennial crop that can achieve high biomass production on marginal lands (Amaducci *et al.*, 2017; Pancaldi & Trindade, 2020; Shepherd *et al.*, 2020; van der Cruijsen *et al.*, 2021). Due to its perennial nature, *Miscanthus* has a limited input requirement and is cultivated under no tillage regime leading to multiple ecosystem services provision (Agostini *et al.*, 2021; Ferrarini *et al.*, 2021; Ferrarini *et al.*, 2017; Martani *et al.*, 2021). Most of the research on *Miscanthus* has been conducted on *Miscanthus x giganteus* (Heaton *et al.*, 2010), which is a naturally occurring sterile triploid hybrid of *Miscanthus sacchariflorus* (*M. sac*) and *Miscanthus sinensis* (*M. sin*) (Hodkinson *et al.*, 2002). Novel *Miscanthus* hybrids (Clifton-Brown, Harfouche, *et al.*, 2019; Clifton-Brown, Schwarz, *et al.*, 2019; Hastings *et al.*, 2017) have been recently obtained from several breeding programs (Clifton-Brown, Schwarz, *et al.*, 2019). In Europe, rhizome- and seed-based *Miscanthus* hybrids are available at a technology readiness levels that can enable the plantation of thousands of hectares per year (Clifton-Brown, Harfouche, *et al.*, 2019). These novel *Miscanthus* hybrids are being tested in multiple environments within the EU-BBI project GRACE.

Plant senescence is a key trait for perennial plants as it limits biomass yield, modifies moisture content and affects nutrient translocation (Boersma *et al.*, 2015; Jensen *et al.*, 2017; Malinowska *et al.*, 2017; Sarath *et al.*, 2014; Yang & Udvardi, 2018). Moisture content at harvest is the most important biomass quality trait (Robson *et al.*, 2013; Styks *et al.*, 2020). Monitoring the dynamics of crop senescence, and moisture content can support the choice of the optimal harvest time that can improve biomass quality and logistics biomass supply chain. Lewandowski *et al.* (2016) found that moisture content of different genotypes can vary due to morphological differences and senescence patterns but it is primarily determined by harvest date. Several studies have shown that late senescence (stay green - SG) maximises biomass yield (Clifton-Brown *et al.*, 2001), while early senescence increases biomass quality (Clifton-Brown & Lewandowski, 2002). SG is determined by a complex physiological control (e.g. chlorophyll efficiency, nitrogen contents, nutrient remobilization and source-sink balance) (Munaiz *et al.*, 2020; Thomas & Howarth, 2000) and traditional phenotyping methods for evaluating SG and delayed senescence are time-consuming (Furbank & Tester, 2011). Non-destructive methods are based on greenness visual score (Bogard *et al.*, 2011) and SPAD measurements (Lopes & Reynolds, 2012; Xie *et al.*, 2016), for the estimation of the green leaf area and relative chlorophyll content respectively. These methods can be used to monitor field trials but are not effective in monitoring senescence dynamics at commercial scale. New sensing technologies have contributed to a substantial improvement in the monitoring of SG in different crops (Cerrudo *et al.*, 2017; Kipp *et al.*, 2014; Liedtke *et al.*, 2020; Lopes & Reynolds, 2012). High-throughput phenotyping (HTP) with remote sensing is a rapid and non-destructive technology that can be used to monitor senescence of

numerous genotypes, thus supporting breeding programs (Anderegg *et al*., 2020; Hassan *et al*., 2018). Remote sensing technologies use different types of sensors, such as Red-Green-Blue (RGB), multispectral, hyperspectral, and thermal sensors, installed on satellites and on unmanned aerial vehicles (UAVs) (Xie & Yang, 2020). Spectral data can be used to calculate vegetation indices (VIs), which can be used to estimate crop traits related to SG trait: Normalized Difference Vegetation Index (NDVI) for green biomass (Cabrera-Bosquet *et al.*, 2011), Enhanced Vegetation Index (EVI) for leaf area index (LAI) (Alexandridis *et al.*, 2020) and Modified Chlorophyll Absorption in Reflectance Index (MCARI) for chlorophyll content (Haboudane *et al.*, 2002). Other VIs, such as the Plant Senescence Reflectance Index (PSRI) (Merzlyak *et al.*, 1999) or the Structure Insensitive Pigment Index (SIPI) (Peñuelas *et al.*, 1995), which are based on the chlorophyll/carotenoid ratio as the decomposition rates of these pigments is affected during senescence, were specifically developed to study crop senescence. The normalized difference water index (NDWI) (Gao, 1996), calculated using near-infrared (NIR) and shortwave-infrared (SWIR) spectral bands, has been proposed as a powerful direct water-sensitive VI, which can be used for the remote sensing of canopy water content (CWC) (Jackson *et al.*, 2004). However, NDWI is rarely calculated by UAV because it requires costly sensors that are equipped with the SWIR band. (Zhang & Zhou, 2019) compared direct against indirect (which does not include the SWIR band) water-sensitive VIs, such as NDVI, NDRE, CIgreen and CIred-edge and found that these VIs were strongly correlated with the CWC as the direct VIs.

Field trials carried out with small plots cannot be monitored using satellite data, for this HTP using UAV-based multispectral images is best used in breeding programs where numerous genotypes are compared (Gracia-Romero *et al.*, 2019; Ostos-Garrido *et al.*, 2019; Su *et al.*, 2019; Varela *et al.*, 2021; Yang *et al.*, 2017; Zhou *et al.*, 2019). UAV-based multispectral images were used in many studies to compare genotypes on the basis of VIs linked to LAI (Potgieter *et al.*, 2017), green LAI (Blancon *et al.*, 2019), canopy cover (Makanza *et al.*, 2018), crop biomass and yield (Johansen *et al.*, 2020; Wang *et al.*, 2019), and senescence dynamics (Hassan *et al.*, 2018). However, many VIs show non-linear relationships with their associated crop traits (Verrelst, Camps-Valls, *et al.*, 2015). Machine learning (ML) regression algorithms have increasingly been used in HTP to recognize non-linear and non-parametric relationships. ML is used to combine multiple VIs for estimating crop traits from a sequence of UAV remote sensing acquisitions. ML models use two main data sets: a training set on which the best model is trained to fit the measured parameters and a test set used to assess the performance of model (Kuhn & Johnson, 2013). In addition to the VIs data, with ML methods numerous types of data, such as categorical variables (e.g. genotype, crop type, locations, agronomic treatments) (Im *et al.*, 2009; Meroni, Waldner, *et al.*, 2021; Wolanin *et al.*, 2020), can be used in the analysis (Verrelst *et al.*, 2019). A ML method commonly used in many remote sensing analyses is

random forest (RF) (Belgiu & Drăguţ, 2016; Holloway & Mengersen, 2018), which can estimate crop biomass (Han *et al.*, 2019) and yield (Johansen *et al.*, 2020) from UAV multispectral images. A main limit of the RF model is its transferability to environments, cropping systems or growing seasons different from those used for training the model (Vuolo *et al.*, 2013). Another limitation is in the training set size (Millard & Richardson, 2015) and the unreliability of predictions made beyond the range of values of the parameters present in the training set (Shah *et al.*, 2019). In addition, Schauberger *et al.* (2020) reported that 52% of the studies on ML do not validate the models' performance with independent test sets. Overall, the quality of training data for developing robust ML models is the key for successfully transferring the trained model and its knowledge to other target domains/tasks. For these reasons, new studies are needed to assess the transferability of ML models for UAV applications in agricultural sciences (Johansen *et al.*, 2020).

However, to date, only time series VIs data from UAV, and not the estimated crop traits of ML models, are used for HTP. A set of known models are normally fitted to VIs time series to characterize plant growth/status associated to different phenological phases. Specifically for the senescence, logistic functions (Christopher *et al.*, 2014) and the Gompertz model (Anderegg *et al.*, 2020) are the two most used models. Another potential approach to fit VIs data is the generalized additive model (GAM) (Nolè *et al.*, 2018). Antonucci *et al.* (2021) for example successfully used GAM approach for HTP of whole-canopy photosynthesis and transpiration. Although remote sensing applications that support these approaches exist and have been already tested successfully for field crops (Alam *et al.*, 2012; Kavats *et al.*, 2019; Yang, 2011; Zhang *et al.*, 2021), no remote sensing application for estimating moisture content of *Miscanthus* are reported in scientific literature.

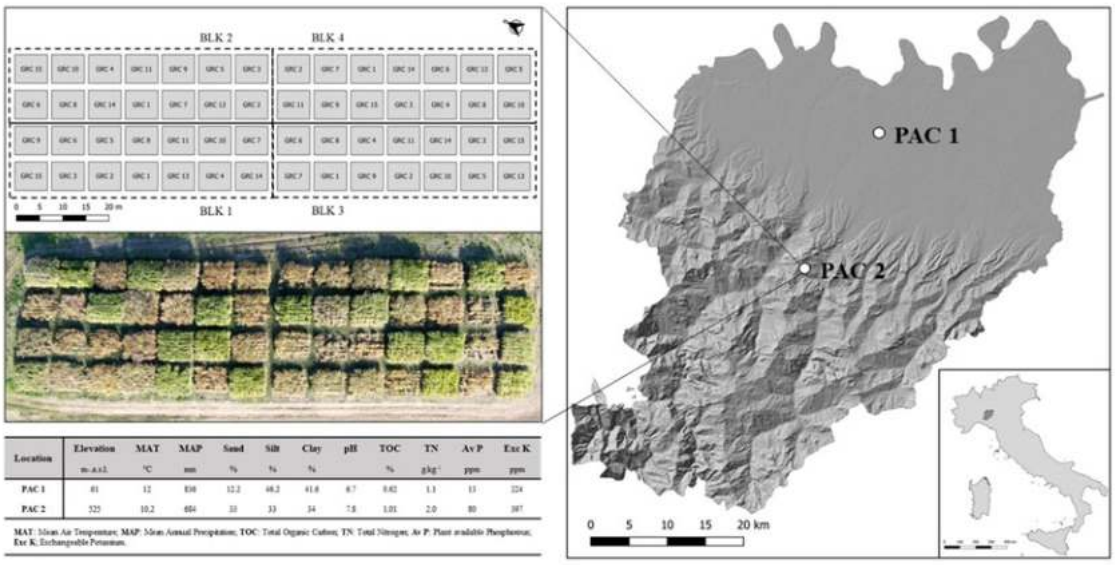
As a first-time testbed for phenotyping *Miscanthus* with UAV remote sensing, two locations, where 14 contrasting *Miscanthus* hybrids were compared in a completely randomized block design, were monitored regularly with moisture content measurements and UAV flights and senescence dynamics were assessed during two growing seasons. The objectives of this study were: 1) to evaluate the performances and transferability of RF models in estimating the moisture content of *Miscanthus* biomass and 2) to phenotype the dynamics of senescence and identify SG trait of contrasting *Miscanthus* hybrids using GAM applied to moisture content time series.

4.2 Materials and methods

4.2.1 Experimental design

This study is part of the EU-BBI funded project GRACE (GRowing Advanced industrial Crops on marginal lands for biorEfineries) that aims to prove the feasibility of large-scale *Miscanthus* cultivation on marginal land. Two of the seven plot scale (PS) trials conducted within GRACE project have been selected for this study. The two sites were located in the province of Piacenza (NW Italy):

PAC 1 located in San Bonico (45°00′11.70″ N, 9°42′35.39″ E) and PAC 2 located in Chiulano (44°50′40.32″ N, 9°35′04.93″ E) (Figure 4.1). Former land use was arable land with cereal crops rotation and permanent meadow respectively, in PAC 1 and PAC 2. The climate in both locations is temperate. The sites differ for soil type and elevation (Figure 4.1). Meteorological data were collected from automatic weather stations located at each experimental site (Table 4.1). Experimental layout was a complete randomized block design with 14 *Miscanthus* hybrids (Table 4.2) with *n*=4 replicates for a total of *n*= 56 plots. Plot size was 6 m x 7 m. The 14 hybrids, coded from GRC 1 to GRC 15 (except GRC 12) were grouped into three main genotypes: *M. x giganteus* as control genotype, interspecific (*M. sin x M. sac*) and intraspecific (*M. sin x M. sin*) hybrids genotypes. Both PS trials were established in April 2018 after winter ploughing and spring seed bed preparation (power harrowing). Plugs and rhizomes were manually transplanted while mechanical weeding during the first years was performed three times. Neither irrigation nor fertilisation were applied. Measurements of this study were carried in the 2nd and 3rd growing season during senescence.



Location	Elevation	MAT	MAP	Sand	Silt	Clay	pH	TOC	TN	Av P	Exc K
	m a.s.l.	°C	mm	%	%	%		%	g kg^{-1}	ppm	ppm
PAC 1	81	12	830	12.2	46.2	41.6	8.7	0.92	1.1	11	224
PAC 2	325	10.2	884	33	33	34	7.8	1.01	2.0	80	307

MAT: Mean Air Temperature; MAP: Mean Annual Precipitation; TOC: Total Organic Carbon; TN: Total Nitrogen; Av P: Plant available Phosphorous; Exc K: Exchangeable Potassium.

Figure 4.1 Locations, experimental field design, main soil properties and drone picture of *Miscanthus* trials.

Table 4.1 Mean monthly temperature, humidity, and rainfall in the two locations considered during the two seasons.

Location	Season	Month	Temperature			Humidity			Rainfall
			Mean (°C)	Max (° C)	Min (° C)	Mean (%)	Max (%)	Min (%)	(mm)
PAC 1	2019	September	19.8	20.2	19.3	76.4	78.5	74.2	79.2
		October	15.5	15.8	15.2	90.1	91.3	88.8	120.6
		November	8.7	8.9	8.5	97.4	97.9	96.9	231.0
		December	4.9	5.2	4.6	91.6	92.6	90.6	73.4
		January	2.9	3.2	2.5	94.1	95.1	93.1	17.8
		February	6.7	7.2	6.2	75.3	77.2	73.3	1.2
	2020	September	20.5	21.0	19.9	73.6	76.1	71.1	18.6
		October	13.1	13.5	12.8	87.4	89.0	85.7	103.8
		November	8.2	8.5	7.9	95.4	96.2	94.6	24.4
		December	3.3	3.5	3.1	99.1	99.3	98.9	216.8
		January	1.5	1.8	1.2	92.5	93.5	91.5	10.4
		February	4.6	4.9	4.3	96.7	97.3	95.9	0.0
PAC 2	2019	September	18.0	18.4	17.7	75.8	77.6	74.0	65.8
		October	14.2	14.5	14.0	85.2	86.7	83.7	120.0
		November	7.7	8.0	7.5	91.0	92.1	89.7	295.6
		December	5.3	5.5	5.0	80.3	81.6	78.9	83.0
		January	5.1	5.4	4.8	77.6	79.7	75.5	34.0
		February	8.3	8.7	7.9	58.2	60.4	56.3	0.6
	2020	September	16.9	17.2	16.5	72.1	73.9	70.1	19.6
		October	11.7	12.0	11.4	82.9	84.7	81.2	102.6
		November	8.3	8.5	8.1	87.5	88.6	86.3	21.4
		December	2.9	3.2	2.7	95.7	96.4	94.9	146.2
		January	2.0	2.3	1.7	80.3	81.9	78.7	94.8
		February	3.9	4.1	3.7	87.3	88.4	86.2	22.0

Table 4.2 Characteristics of the 14 *Miscanthus* hybrids considered in this study.

Material	Hybrid code	Genotype	Planting density
Seed-based plugs	GRC 1-8	*M.sinensis x M. sinensis*	3 plants/m^2
Rhizomes *	GRC 9	*M. x giganteus*	1.5 plants/m^2
Seed-based plugs	GRC 10 - 14 (except GRC 12)	*M.sinensis x M. sacchariflorus*	1.5 plants/m^2
Rhizomes *	GRC 15	*M.sinensis x M. sacchariflorus*	1.5 plants/m^2

* Hybrids commercially available

4.2.2 Crop measurements

Senescence was tracked visually following the scoring method proposed by Robson *et al.* (2013), which is based on a scale from 1 to 9, where 1 indicates the lowest level of "greenness" of the whole visible aerial parts of the plant and 9 is the score attributed when no visible leaf senescence occurs. Scores were acquired from August to February (until harvest) for a total of 10 events in PAC 1 and 9 in PAC 2. Beside scoring senescence, at each measurement event whole stems samples randomly selected for each plot (20 for *M. sin x M. sin* and 10 for *M. sin x M. sac* hybrids respectively) were

sampled to calculate plant moisture content. Samples were weighed immediately after harvest and again after having been oven dried at 105°C and then percentage of moisture content was calculated (Samuelsson *et al.*, 2006).

4.2.3 UAV multispectral data and vegetation indices

The unmanned aerial vehicle (UAV) used in the experiment was a four-rotator DJI Matrice 210 RTK (SZ DJI Technology Co., Shenzhen, Guangzhou, China) combined with a RTK (Real Time Kinematic) GPS positioning system. At each visual scoring event a UAV multispectral data acquisition was performed, in addition 10 supplementary flight missions were carried out on PAC 1 and 5 on PAC 2 to increase the frequency of senescence tracking. Ten flights were performed over PAC 1 in both seasons, while in PAC 2, 6 and 8 flights were realized in the first and second seasons, respectively (Table 4.3). The UAV was equipped with a MicaSense RedEdge-Mx multispectral camera (MicaSense, Seattle, WA, USA). RedEdge-Mx camera can acquire 5 different spectral bands images: blue (475 nm centre, 32 nm FWHM), green (560 nm centre, 27 nm FWHM), red (668 nm centre, 14 nm FWHM), red edge (717 nm centre, 12 nm FWHM) and near-infrared (840 nm centre, 57 nm FWHM).

Table 4.3 Unmanned Aerial Vehicle (UAV) flights performed in the two locations along two senescence seasons.

Location	Season	UAV flights days
PAC 1	2019	05/08/2019, 10/09/2019, 27/09/2019, 11/10/2019, 26/10/2019, 09/11/2019, 20/11/2019, 03/12/2019, 10/01/2020, 21/02/2020
	2020	01/09/2020, 17/09/2020, 28/09/2020, 10/10/2020, 28/10/2020, 06/11/2020, 19/11/2020, 15/12/2020, 26/01/2021, 01/02/2021
PAC 2	2019	12/09/2019, 12/10/2019, 09/11/2019, 10/12/2019, 10/01/2020, 14/02/2020
	2020	16/09/2020, 29/09/2020, 10/10/2020, 28/10/2020, 06/11/2020, 23/11/2020, 13/12/2020, 19/02/2021

All flights were performed between 11.00 and 15.00. The flight altitude above ground level (AGL) was 40-50 m in PAC 1 and 80-100 m in PAC 2. The forward overlap was set at 80% and lateral overlap was set at 75% of the images. The flight speed was set at 3 m s^{-1}. The ground sampling distance (GSD) was 2.78-3.47 cm and 5.56-6.94 cm in PAC 1 and PAC 2 respectively. The flight was performed in automatic mode with waypoints routes as the presence of a GPS navigation system enables a more accurate image acquisition. The DJI Pilot software (SZ DJI Technology Co., Shenzhen, Guangzhou, China) was used for flight planning and automatic mission control. For the

radiometric calibration of the images, the reflectance of a spectral panel (MicaSense, Seattle, WA, USA) with reflectance values provided by MicaSense, was captured before each flight. In addition, a light sensor that automatically adjusts the readings to ambient light was mounted at the top of the UAV to minimize error during image capture. The radiometric calibration, image mosaicking and orthomosaic generation were done using the Pix4D mapper (Pix4D, S.A., Lausanne, Switzerland). The orthomosaic in reflectance values generated from the software was used for the calculation of 54 vegetation indices (VIs) as shown in Table 4.4.

Table 4.4 List of the vegetation indices evaluated in random forest models for the estimation of *Miscanthus* moisture content.

VIs	Equation	Reference	VIs	Equation	Reference
BNDVI	(NIR-Blue)/(NIR+Blue)	Wang *et al.*, 2007	MTVI1	1.2*(1.2*(NIR-Green)-2.5*(Red-Green))	Haboudane *et al.*, 2004
Chlrededge	RedEdge/Red - 1	Gitelson *et al.*, 2006	MTVI2	1.5*((1.2*(NIR-Green)-2.5*(Red-Green))/(sqrt((2*NIR+1)^2-(6*NIR-5*sqrt(Red)-0.5))))	Haboudane *et al.*, 2004
CIgreen	NIR/Green-1	Gitelson *et al.*, 2003	NDRE	(NIR-RedEdge)/(NIR+RedEdge)	Gitelson & Merzlyak, 1994
CIre	NIR/RedEdge-1	Gitelson *et al.*, 2003	NDVI	(NIR-Red)/(NIR+Red)	Rouse *et al.*, 1973
Datt1	(NIR-RedEdge)/(NIR+Red)	Datt, 1999	NGBDI	(Green-Red)/(Green+Blue)	Wang *et al.*, 2007
Datt2	NIR/RedEdge	Datt, 1999	NGRDI	(Green-Red)/(Green+Red)	Tucker, 1979
DVI	NIR - Red	Richardson & Wiegand, 1977	NLI	(NIR^2-Red)/(NIR^2+Red)	Chen, 1996
EVI	2.5*((NIR-Red)/(1+NIR+6*Red-7.5*Blue))	Huete *et al.*, 2002	OSAVI	(1+0.16)*(NIR-Red)/(NIR+Red+0.16)	Rondeaux *et al.*, 1996
EVI2	2.4*((NIR-Red)/(1+NIR+Red))	Miura *et al.*, 2008	OSAVI2	(1+0.16)*(NIR-RedEdge)/(NIR+RedEdge+0.16)	Wu *et al.*, 2008
EVI3	2.5*((NIR-Red)/(NIR+2.4*Red+1))	Jiang *et al.*, 2008	PNDVI	(NIR-Blue+Green+Red)/(NIR+Blue+Green+Red)	Wang *et al.*, 2007
GARI	(NIR-(Green-(Blue-Red)))/(NIR-(Green+(Blue-Red)))	Gitelson *et al.*, 1996	PSRI	(Red-Blue)/RedEdge	Merzlyak *et al.*, 1999
GBNDVI	(NIR-Green+Blue)/(NIR+Green+Blue)	Wang *et al.*, 2007	RBNDVI	(NIR-Red+Blue)/(NIR+Red+Blue)	Wang *et al.*, 2007
GDVI	NIR-Green	Tucker, 1979	RDVI	(NIR-Red)/((NIR+Red)^0.5)	Broge & Leblanc, 2001
GI	Green/Red	Smith *et al.*, 1995	rededgeWDRVI	(0.1*NIR-RedEdge)/(0.1*NIR+RedEdge)+(1-0.1)/(1+0.1)	Gitelson, 2004
GLI	(2*Green -Red - Blue)/(2*Green+Red+Blue)	Widlowski *et al.*, 2000	RI	(Red-Green)/(Red+Green)	Escadafal & Huete, 1991
GNDVI	(NIR-Green)/(NIR+Green)	Gitelson *et al.*, 2006	SAVI	((NIR-Red)/(NIR+Red+0.5))*(1+0.5)	Huete, 1988
GOSAVI	(NIR-Green)/(NIR+Green+0.16)	Sripada *et al.*, 2006	SIPI	(NIR - Blue) / (NIR - Red)	Peñuelas *et al.*, 1995
greenWDRVI	(0.1*NIR-Green)/(0.1*NIR+Green)+(1-0.1)/(1+0.1)	Gitelson, 2004	SR	NIR/Red	Birth & McVey, 1968
GRNDVI	(NIR-Green+Red)/(NIR(Green+Red)	Wang *et al.*, 2007	TCI	1.2*(RedEdge-Green)-1.5*(Red-Green)*sqrt(RedEdge/Red)	Hunt *et al.*, 2011
GRVI	NIR/Green	Tucker, 1979	TNDVI	sqrt(0.5 + (NIR-Red)/(NIR+Red))	Bannari *et al.*, 2002
IPVI	((NIR/(NIR+Red))/2)*((NIR-Red)/(NIR+Red)+1)	Crippen, 1990	TRBI	(Green+Red)/NIR	Vincini & Frazzi, 2011
MCARI	((RedEdge-Red)-0.2*(RedEdge-Green))*(RedEdge/Red)	Daughtry *et al.*, 2000	VARIgreen	(Green-Red)/(Green+Red-Blue)	Gitelson *et al.*, 2002
MCARI/MTVI2	MCARI/MTVI2	Eitel *et al.*, 2007	VARIrededge	(RedEdge-Red)/(RedEdge+Red)	Gitelson *et al.*, 2002
MCARI/OSAVI	MCARI/OSAVI	Wu *et al.*, 2008	WDRVI	(0.1*NIR-Red)/(0.1*NIR+Red)+(1-0.1)/(1+0.1)	Gitelson, 2004
MCARI/OSAVI2	MCARI/OSAVI2	Wu *et al.*, 2008	WDRVI2	(0.2*NIR-Red)/(0.2*NIR+Red)+(1-0.2)/(1+0.2)	
MCARI2	((NIR-RedEdge)-0.2*(NIR-Green))*(NIR/RedEdge)	Wu *et al.*, 2008	WDRVI3	(0.1*NIR-Red)/(0.1*NIR+Red)	
MSAVI	((2*NIR+1-sqrt((2*NIR + 1)**2-(8*(NIR-Red))))/2)	Qi *et al.*, 1994	WDRVI4	(((0.1*NIR-Red)/(0.1*NIR+Red))+1)/2	

To extract the spectral information of each experimental plot, the polygons of the experimental design were drafted in AutoCAD (Autodesk, San Rafael, California, USA) and georeferenced based on the UAV multispectral images by using QGIS software (QGIS Development Team, 2021).

4.2.4 Machine learning model for moisture content estimation

A recursive feature elimination (RFE) algorithm (Feng *et al.*, 2020; Yue *et al.*, 2018) was initially applied to solve the multi-collinearity problem among VIs by selecting the most important VIs for moisture content estimation. Inputs for the RFE algorithm were the predictors variables (the 54 VIs calculated from UAV multispectral images) and the corresponding target values (the measured plant moisture content). In the RFE algorithm, the random forest (RF) model was used to minimize the root mean square error (RMSE). The RFE results were combined with the *pickSizeTolerance* function to select a model containing fewer predictors variables within the bounds of a user-defined threshold metric (Parmley *et al.*, 2019). RMSE metric and the 0%, 1% and 5% tolerance thresholds were utilized to identify models with acceptable performance but with fewer predictors variables.

On the selected VIs, RF was then used to estimate the moisture content of *Miscanthus* hybrids. RF model is an ensemble learning model where the output averages the result of multiple regressions trees (Kamir *et al.*, 2020). The RF models were created using the *caret* R package (Kuhn, 2008). Two steps in RF modelling were adopted: firstly, RF was trained and tested on the VIs selected from RFE algorithm at the tolerance threshold of 1%; secondly, the three categorical crop variables (material, hybrid code, and genotype, Table 4.2) and their combinations were added in RF modelling to check for improvement in model's performance.

For the RF modelling, the optimal size of the variable subset (*mtry*) was obtained by grid-searching method using repeated k-fold cross-validation. The repeated *k*-fold cross-validation consist of dividing the data into *k* independent folds of the same size, training the algorithm on (*k*-1) folds and testing its accuracy on the remaining fold based on the error between predicted and target values several times (Kamir *et al.*, 2020). In our study we used a ten-fold cross validation, which was repeated 5 times. This procedure was used to estimate the moisture content and to evaluate the transferability of the models tested on 5 subset test datasets: 4 specific season- and location-datasets (two locations x two growing seasons) and one reference dataset, as a comparison. The reference dataset was created by using a stratified random sampling method (Han *et al.*, 2019): data from both locations and seasons were split into 70/30 between training and testing based on moisture content distribution. To include the categorical variables into the models (second step), a one-hot-encoded approach was used to encode categorical variables into numbers, assigning the value 1 when the condition is satisfied and 0 when it is not satisfied.

RF models' performances were evaluated calculating the root mean square error (RMSE, Equation 4.1) and the normalized root mean square error (NRMSE, Equation 4.2) as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_i - y_i)^2}{n}} \quad \text{Equation 4.1}$$

$$NRMSE\ (\%) = \frac{\sqrt{\frac{\sum_{i=1}^{n}(x_i - y_i)^2}{n}}}{\bar{y}} 100 \quad \text{Equation 4.2}$$

where n is the sample number, x_i and y_i are the estimated and measured moisture content, and $\bar{y}$ is the mean of the measured value. The performance metrics were also calculated for different intervals of moisture content and each *Miscanthus* hybrid. The moisture content intervals investigated were lower than 30%, between 30% and 60%, higher than 60%, and finally between 10% and 80%. The set size used for each training dataset was reported to compare the metrics of the models. For each model, the RMSE and NRMSE were calculated for each genotype and for the different moisture content intervals to evaluate the models.

4.2.5 GAM for phenotyping *Miscanthus* senescence dynamics

The moisture content during senescence was estimated from spectral data acquired by UAV using the validated RF model: this approach was selected to add supplementary flights to the dataset without field measurements. The validated RF model included as predictors variables the VIs and the three categorical variables. The time series moisture content dataset estimated from RF was fitted against the modified days of the year (DOY). Early and late senescence in *Miscanthus* occur normally in two different years. To overcome the problem of having non-sequential DOY data along the senescence season, moisture content data of January and February were calculated by adding 365 to the DOY of the corresponding year. To phenotype the dynamics of senescence and identify stay-green (SG) trait of the different *Miscanthus* hybrids, statistical analysis of the estimated moisture content time series was carried out via a generalized additive model (GAM). The regression model GAM is a non-parametric extension of the generalized linear model (GLM), which allows the integration of non-parametric smoothing functions and non-linear fitting of the variables. GAM models were fitted in R package *mgcv* (Wood, 2017). The fitted model used fixed factors and a smooth for DOY, based on location, season, and hybrid. GRC 9 (*M. x giganteus*) was used a reference to detect difference between interspecies and intraspecies *Miscanthus* hybrids.

4.3 Results

4.3.1 Dynamics of moisture content in *Miscanthus* biomass

The variation of moisture content and its frequency distribution measured during the two senescence seasons at two locations is shown in Figure 4.2 and in Figure 4.3. Overall, moisture content loss started at the beginning of December at both locations and for all genotypes (Figure 4.2). *M. sin x M. sac* hybrids showed a higher moisture content (+ 18 % and + 6 %) than *M. sin x M. sin* hybrids and *M. x giganteus* in both locations from December until harvest in late winter (Figure 4.2 and Figure 4.3). On average, *M. sin x M. sac* hybrids and the *M. x giganteus* were harvested at 45 % and 37 % moisture content respectively (Figure 4.2 and Figure 4.3). *M. sin x M. sin* hybrids had an average moisture content at winter harvest of 22%. The dynamics of moisture content during senescence is confirmed by visual recording of senescence score based on plant greenness (Figure 4.4). For all genotypes, the correlation between senescence score and moisture content indicated that moisture content loss starts when senescence score values of 4 are recorded.

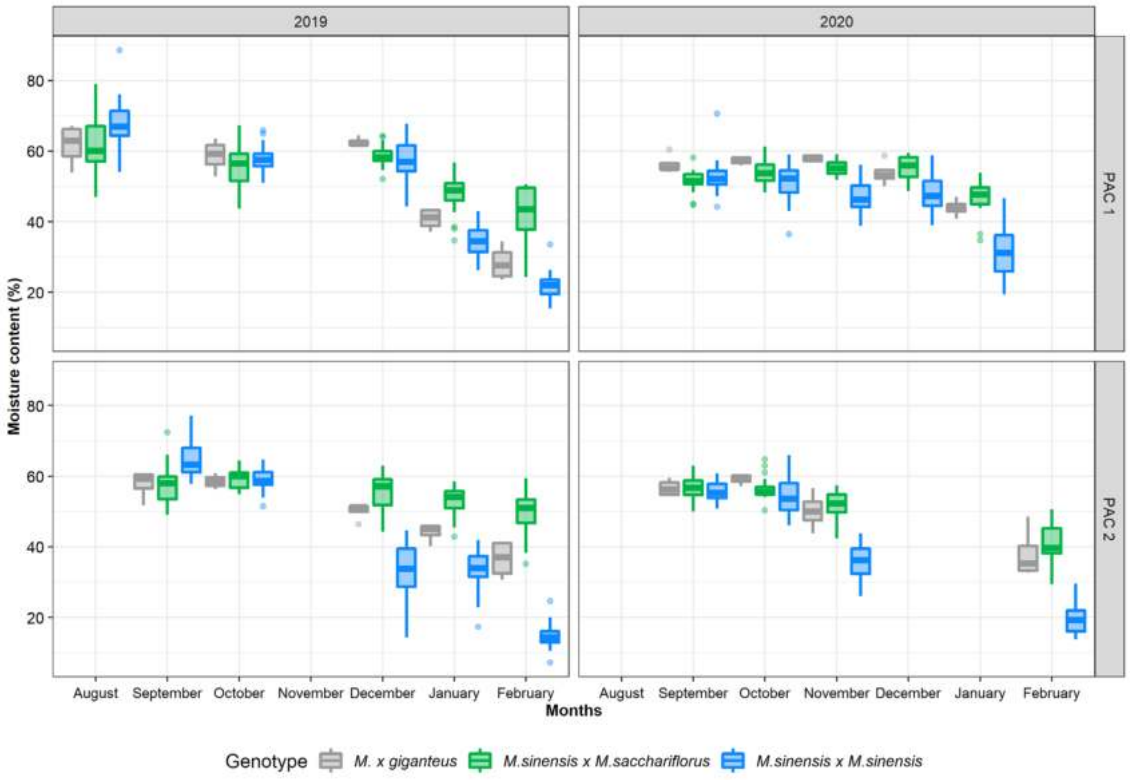


Figure 4.2 Temporal dynamic of the measured moisture content of different *Miscanthus* genotypes along two growing seasons and two locations.

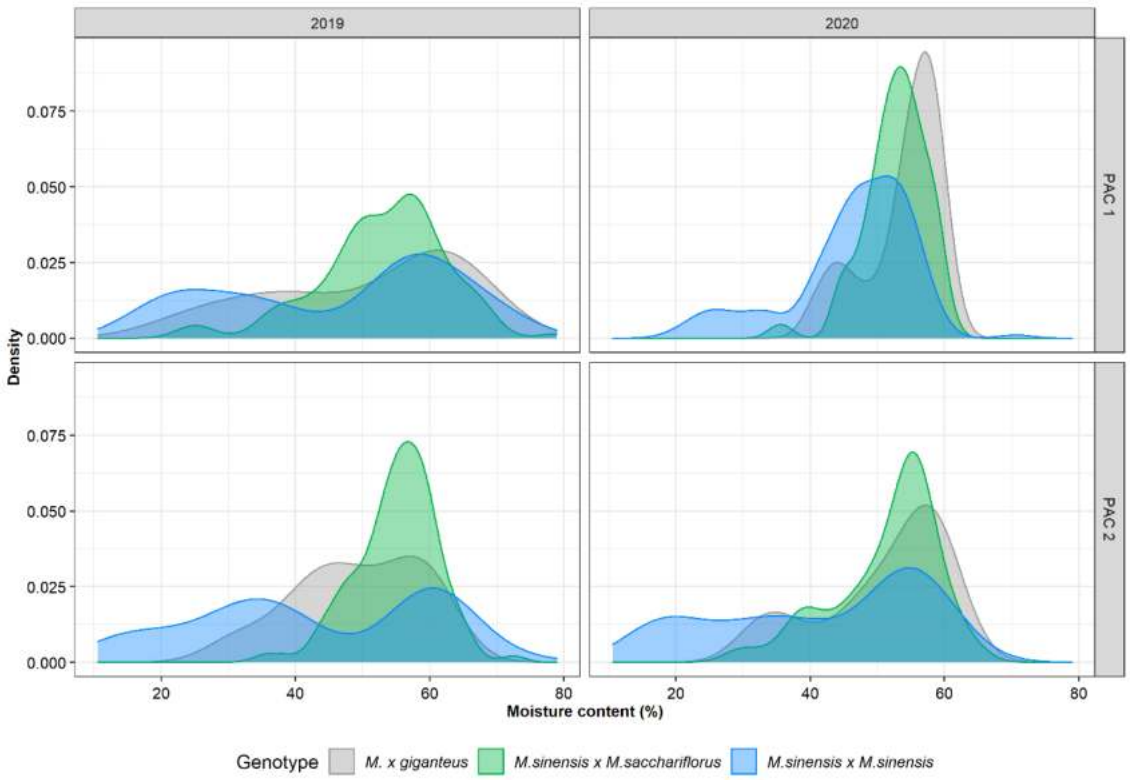


Figure 4.3 Frequency distribution of the moisture content of different *Miscanthus* genotypes during the two seasons and on two locations.

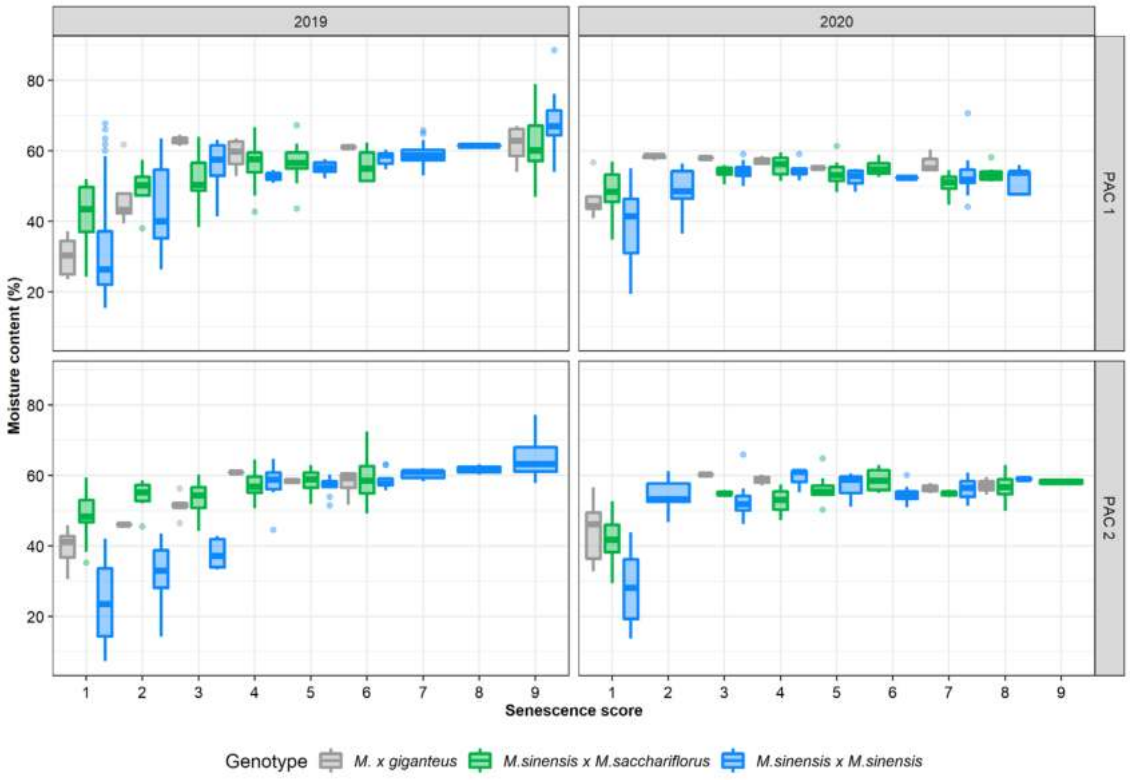


Figure 4.4 Relation between senescence score and moisture content of different *Miscanthus* genotypes.

4.3.2 Recursive Feature Elimination of vegetation indices

The optimal number of vegetation indices (VIs) included in the models to minimize RMSE in the estimation of moisture content was obtained by the recursive feature elimination (RFE) algorithm with repeated cross-validation (Figure 4.5). RFE analysis showed that using 4 or less VIs led to a moisture content estimation with RMSE values higher than 8 % (Figure 4.5a). With the 0% tolerance threshold, the minimum RMSE (7.4%) was achieved with 30 VIs. However, the use of 20 or more VIs led to a moisture content estimation with a mean RMSE value of 7.4%. On the contrary, with the thresholds of tolerance of 1% and 5%, the optimal number of VIs was 14 (RMSE = 7.5 %) and 6 (RMSE = 7.8%) respectively (Figure 4.5a). The threshold of tolerance of 1% was chosen as the threshold that maximises the model's performances with the minimum number of VIs. According to the importance of the ranking (Figure 4.5b), 14 VIs have been selected for RF models training among the 54 VIs calculated (Figure 4.5b). The 14 VIs were: BNDVI, GDVI, PSRI, MCARI/MTVI2, GOSAVI, NGBDI, NLI, GBNDVI, GLI, MCARI/OSAVI2, SIPI, MCARI2, OSAVI2 and GI. The 6 most important VIs to reach 5% tolerance (RMSE < 7.8 %) were (Figure 4.5b): BNDVI, GDVI, PSRI, MCARI/MTVI2, GOSAVI and NGBDI.

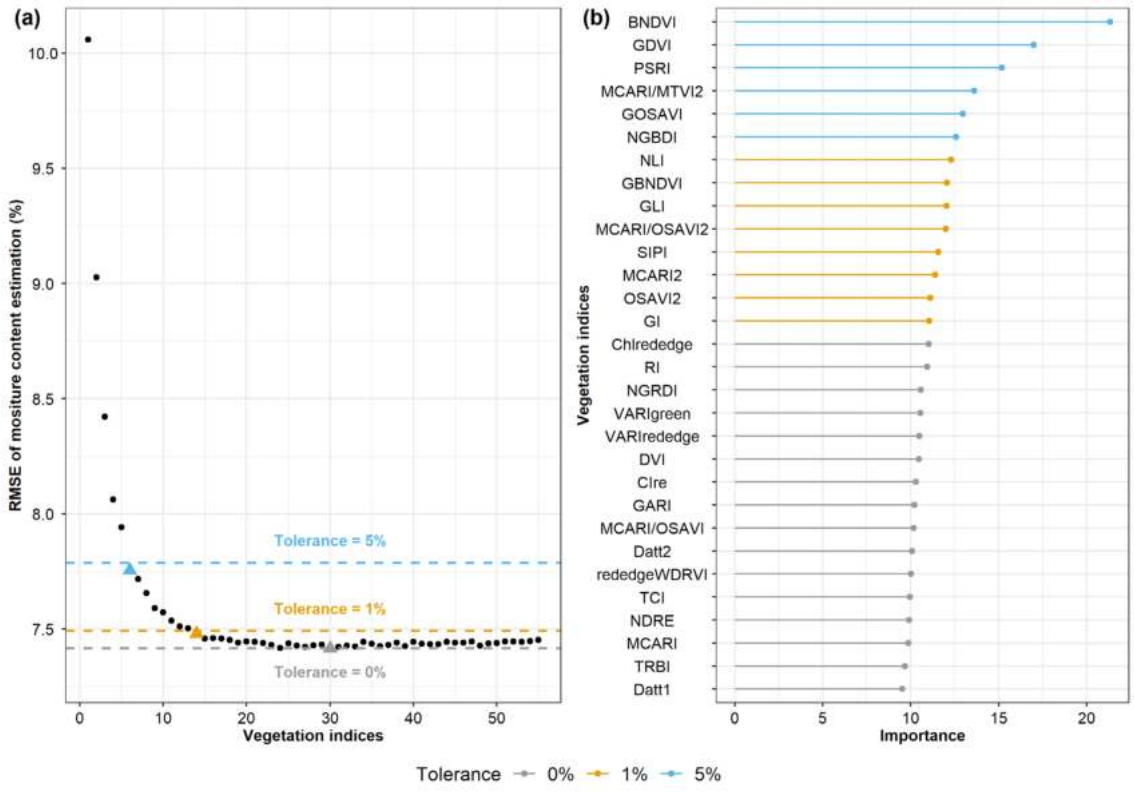


Figure 4.5 (a) Results of the RFE algorithm with different tolerance thresholds and (b) importance of the VIs used in the different tolerance thresholds (Blue = 5%, Yellow =1% and Grey = 0%).

4.3.3 RF model performance and transferability

The performances (RMSE and NRMSE) of the random forest (RF) models were compared among the season-specific datasets of the two location and against one reference dataset (split into 70/30 training/test) (Table 4.5). When all the genotypes and all moisture content intervals were considered, the RF model of the reference dataset was the most accurate one among the 5 models considered in estimating *Miscanthus* moisture content (RMSE = 6.9% and NRMSE = 14%). The other models achieved lower accuracy values with RMSE ranging from 9.2% to 10.6% and NRMSE from 20.1% to 22.1%. The accuracy of the RF models trained with the season-location specific datasets and for the intervals of moisture content of 30-60% and > 60% was on average similar (RMSE =8.5%) to the accuracy of the RF model trained with the reference dataset for the same intervals (RMSE = 6.3%). On the contrary, the accuracy of the RF models for the season-location specific datasets was lower for the interval of moisture content < 30% (RMSE =16.4%) than the reference dataset (RMSE =10.7% and 6.3% respectively).

Table 4.5 Results of the RF models performance according to location, season, moisture content intervals and genotype (blue = *M. sin x M. sin*, grey = *M. x giganteus* and green = *M. sin x M. sac*).

Test dataset VIs		Intervals of moisture content (%)	*M. sin x M. sin*			*M. x giganteus*			*M. sin x M. sac*			All Genotypes		
Location	Season		n	RMSE	NRMSE	n	RMSE	NRMSE	n	RMSE	NRMSE	n	RMSE	NRMSE
PAC 1	2019	x < 30	81	19.7	87.5	0	18.4	75.9	2	21.0	83.8	83	19.7	86.7
		30 ≤ x ≤ 60	296	6.1	12.5	49	5.9	13.1	249	6.4	12.5	594	6.2	12.5
		x > 60	45	11.9	18.1	6	10.0	15.8	26	9.4	14.4	77	11.0	16.9
		10 ≤ x ≤ 80	422	12.2	26.2	55	9.6	19.0	277	7.9	14.9	754	10.6	21.6
	2020	x < 30	105	4.5	18.2	2			5			112	4.5	18.2
		30 ≤ x ≤ 60	229	8.8	18.2	40	11.8	22.1	228	11.6	22.0	497	10.2	20.2
		x > 60	83	12.2	17.2	13	4.7	7.7	44	6.0	9.7	140	8.3	12.9
		10 ≤ x ≤ 80	417	8.5	18.4	55	11.6	21.5	277	11.5	21.9	749	9.9	20.2
PAC 2	2019	x < 30	84	22.7	113.2	2			5			91	22.7	113.2
		30 ≤ x ≤ 60	303	9.0	20.7	43	4.6	9.8	247	5.2	9.8	593	7.0	14.5
		x > 60	48	8.4	13.3	11	4.6	7.7	25	6.3	10.0	84	7.6	12.1
		10 ≤ x ≤ 80	435	13.6	31.7	56	4.6	9.4	277	5.4	9.9	768	10.6	22.1
	2020	x <30	84	19.2	93.0	2			3	9.4	31.8	89	18.8	88.9
		30 ≤ x ≤ 60	288	5.6	11.6	45	9.2	18.5	254	4.8	9.3	587	5.7	11.4
		x > 60	76	5.8	9.4	12	3.5	5.8	40	6.7	10.7	128	5.9	9.5
		10 ≤ x ≤ 80	448	11.1	26.6	59	8.7	17.0	297	5.1	9.9	804	9.2	20.1
Reference dataset*		x < 30	89	10.8	51.8	1	10.8	43.3	3	8.9	32.3	93	10.7	50.0
		30 ≤ x ≤ 60	260	5.1	10.5	41	7.8	15.7	227	7.1	13.5	528	6.3	12.5
		x > 60	61	7.3	11.6	10	5.0	8.1	31	4.7	7.4	102	6.3	10.0
		10 ≤ x ≤ 80	410	6.8	14.9	52	7.6	14.9	261	6.9	12.9	723	6.9	14.0

Abbreviations: **n** is the training set size, **RMSE**: root mean square error (%), **NRMSE**: normalized root mean square error (%).

*** Reference dataset** is composed by 30 % of the initial dataset (both locations and seasons) used for the validation and 70% used for RF model training.

The addition of categorical variables (material, hybrid code and genotype of Table 4.2) to the reference dataset model of VIs improved the accuracy of moisture content estimation (Figure 4.6). The single addition of material, hybrid code or genotype in the model (Figure 4.6b, c, d) decreased the RMSE from 6.9% (model with only Vis) to 6.8%, 6.4% and 5.7 % respectively. The simultaneous addition of three categorical variables to the model, achieved the best performance with an RMSE = 5.6% and NRMSE= 11.4% (Figure 4.6e). Finally, the RMSE of all models was evaluated for each genotype (Figure 4.6f). The addition of categorical variables decreased the RMSE value with respect to the model with only VIs for the *M. x giganteus* genotype from 7.6% to 5.6%, for the interspecific *M. sin x M. sac* genotype hybrids from 6.9% to 4.7%, while for intraspecific *M. sin x M. sin* genotype hybrids from 6.8% to 6.1%.

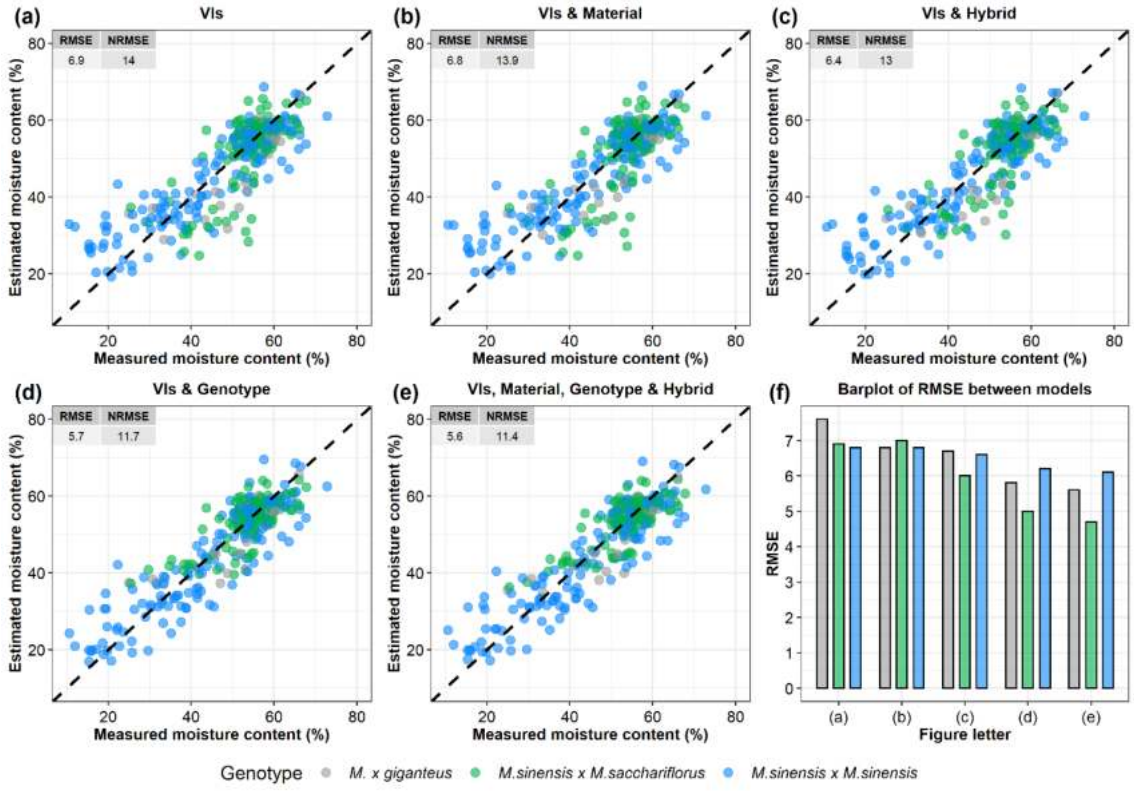


Figure 4.6 Estimated vs measured moisture content (%) of *Miscanthus* with RF model with only VIs and no categorical variables (a), with the addition of transplanting material (b), of hybrid (c), of genotype (d) and their combination €. RMSE for each model are reported as barplot (*f*) according to the different genotype.

4.3.4 Phenotyping of *Miscanthus* senescence dynamics with multiple UAV flights

The RF model trained with the VIs and the three categorical variables was used to estimate moisture content of *Miscanthus* hybrids from spectral data of multiple UAV flights at two locations. Generalized additive model (GAM) was applied to time series moisture content data estimated from RF model, with the *M. x giganteus* (GRC 9) as reference for estimating significant differences among

the hybrids during senescence. *M. sin x M. sin* hybrids (GRC 1-8) from DOY 280 (mid-early October) showed a constant and significant lower moisture content than the *M. x giganteus* hybrid (Figure 4.7). The first genotype showing a significant difference in moisture content compared to GRC 9 was GRC 5, at DOY 260 (mid-September), while the last was GRC 1, at DOY 312 (mid-early November). Intraspecies *M. sin x sin* hybrids showed the highest variability on moisture content loss during senescence compared to interspecies *M. sin x M. sac* hybrids. The estimated difference of moisture content at harvest varied from 10.2% for GRC 1 to 14.5% for GRC 6. On the contrary, constant negative differences compared to GRC9 occurred later in the season (early November) for interspecific *M. sin x M. sac* hybrids (GRC 10-15). The difference is statistically significant approximately from DOY 295 (mid-late October) for GRC 10 hybrid and from DOY 314 (mid-early November) for GRC 13 hybrid. At harvest, the estimated moisture content difference varied from -9.2% for GRC 11 to -10% for GRC 14. The rhizome-based GRC 15 hybrid, a *M. sin x M. sac* genotype, showed a similar moisture content dynamics to the other rhizome-based hybrid (GRC 9).

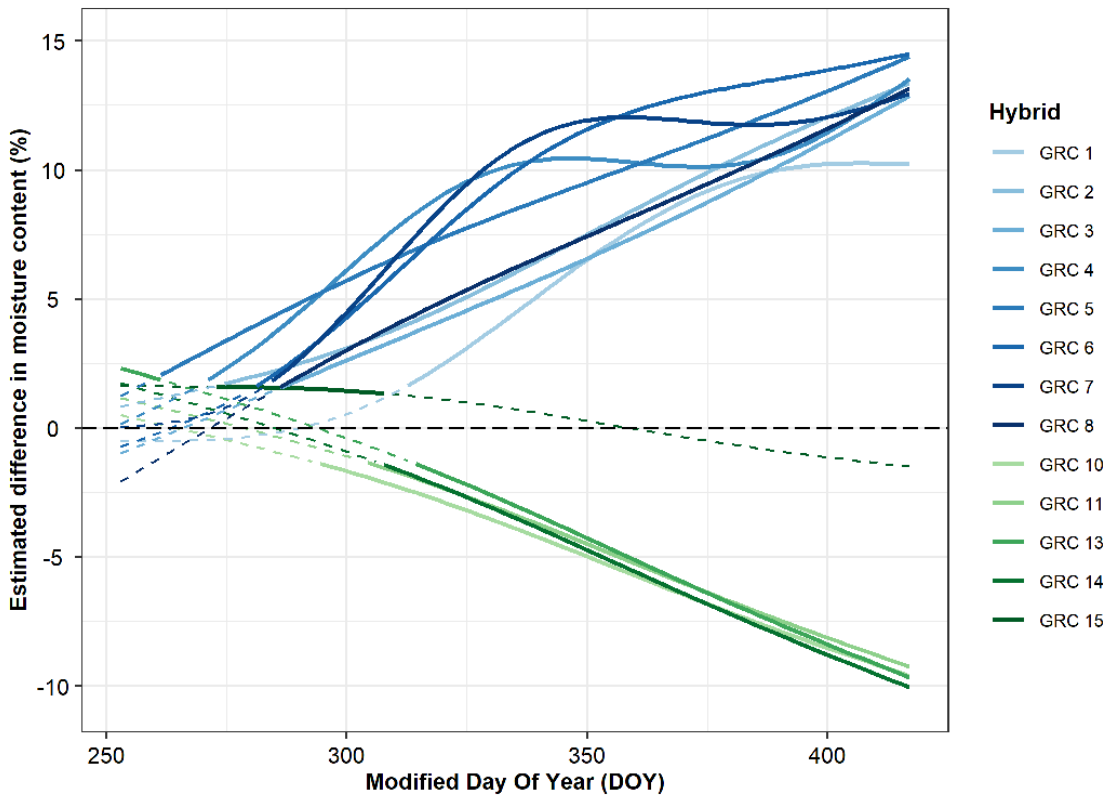


Figure 4.7 Senescence dynamics of the different *Miscanthus* hybrids according to the difference in estimated moisture content with reference hybrid *M. x giganteus* - GRC 9 (dashed black line). The estimation of moisture content time series was carried out by using a GAM. Solid and dashed coloured lines denotes respectively significant ($P<0.05$) and not significant differences of the corresponding hybrid compared to reference hybrid.

4.4 Discussion

The characterization of moisture content dynamics of *Miscanthus* biomass is important to determine the harvest time and selecting the most suitable genotypes in each environment. This study estimated the moisture content of 14 contrasting *Miscanthus* hybrids combining unmanned aerial vehicle (UAV) remote sensing and machine learning. The random forest (RF) model was trained with moisture content values measured directly from each plot trial, UAV multispectral data (the vegetation indices) and categorical variables of *Miscanthus* hybrids (material, hybrid code and genotype). The time series of the moisture content values estimated by RF model from VIs derived from multiple UAV flights were used for phenotyping senescence dynamics and identifying the stay-green (SG) trait of *Miscanthus* hybrids using the generalized additive model (GAM).

4.4.1 Selection of multispectral vegetation indices for *Miscanthus* moisture content estimation

Increasing the number of VIs from 1 to 14 improved the RF model's accuracy and allowed to decrease RMSE from 10 % to 7.5 % (Figure 4.5a). Generally, the estimation of the crop traits via multiple VIs is affected by data redundancy and multi-collinearity among some vegetation indices (VIs) (Yue *et al.*, 2018). The use of recursive feature elimination (RFE) algorithm proved to be a suitable approach to minimise RMSE while reducing the noise effect caused by data redundancy and multi-collinearity, as suggested by Anderegg *et al.* (2020) and Han *et al.* (2019). This study showed that the three most important VIs for estimating moisture content were VIs based on blue (BNDVI), green (GDVI) and red-edge (PSRI) spectral bands (Figure 4.5b). Zhu *et al.* (2019) found that the blue band is sensitive to the change of carotenoid content and the green and red-edge bands are sensitive to the change of chlorophyll content. VIs based on these spectral bands indeed have been used to study crop senescence dynamics (Anderegg *et al.*, 2020; Peñuelas & Inoue, 1999). The blue band proved to be the most important variable for predicting harvest date (pod's maturity) in soybean (Yu *et al.*, 2016). Anderegg *et al.* (2020) reported that the time series of PSRI could accurately track senescence dynamics of the canopy of wheat and replace the visual scorings. Furthermore, the SIPI was strongly correlated with relative water content (RWC) and can indirectly evaluate leaf water stress (Peñuelas & Inoue, 1999). Also, this study confirmed that the VIs selected by the RFE algorithm and used in the RF model were sensitive to changes of chlorophyll/carotenoid ratio during senescence. Finally, although no VIs based on the SWIR band were used in this study, it was demonstrated that the combination of multiple VIs based on VIS-NIR images compensated for the lack of the SWIR band, which is known to predict well crop moisture content when integrated with VIs such as NDWI (Zhang & Zhou, 2019).

4.4.2 Moisture content estimation with a machine learning algorithm

This study estimated the moisture content with the RF model, trained with a wide range of genotypes, across two senescence seasons and at two different locations, differing strongly in soils and slightly in climate. These differences, as suggested by Maxwell *et al.* (2018), help to assess the RF model transferability. The transferability of the moisture content estimation models was evaluated by splitting the moisture content dataset into 5 test datasets. The performance metrics of the RF models showed that a good accuracy (6.9% of RMSE and 14.0% of NRMSE) was achieved when all the genotypes and all moisture content intervals were considered in the models (Table 4.5). Similar results were reported by Li *et al.* (2021) to estimate the moisture content of three species of trees, who achieved a NRMSE between 8.6% and 13.9%. The models evaluated to estimate the moisture content might be affected by errors in the estimation in some moisture content intervals due to limits in the range of data used to train the model (Shah *et al.*, 2019). Indeed, small increases in RF models performance were found when the models were trained with the specific season- and location-datasets. This difference is due to different models' accuracy when the moisture content is < 30%. During the two seasons, many hybrids did not reach such low moisture content, and thus the training set size for this interval was lower.

To assess the performance of the models in identifying the optimal harvest dates based on moisture content at different endpoints of drying, the moisture content dataset was indeed divided into different intervals (<30%, 30-60, >60% and 10-80%). It is considered that the optimal moisture content for the *Miscanthus* winter harvest is at or below 20% (Lewandowski *et al.*, 2016) in order to avoid self-ignition of biomass, minimise transport costs and an increase combustion efficiency (Robson *et al.*, 2013). In this study, especially novel interspecies seed-based *M. sin x M. sac* hybrids rarely reached at harvest a moisture content lower than 30% (Figure 4.2), while *M. sin x M. sin* in some cases dried until 10 %. In the low moisture content interval (<30%), a large difference in RMSE was found between the model trained with the reference dataset and on the season-location specific datasets (Table 4.5). These results indicate that the tested models cannot be transferred with good accuracy to locations and or/growing seasons where biomass of these genotypes dried until moisture content <30%. The low transferability of RF beyond the extreme values of the training data range confirmed that this is one of the main limits of the RF model (Johansen *et al.*, 2020; Vuolo *et al.*, 2013). On the contrary, the RF models were transferable in different locations and growing seasons for moisture content values ranging between 30 and 60% (Table 4.5). The training set size and the moisture content distribution during senescence confirmed to be the most important dataset's characteristics to achieve good model's performances (Millard & Richardson, 2015) and transferability (Johansen *et al.*, 2020).

The addition of categorical variables in RF model improved the estimation of moisture content. Introducing three categorical variables such as material, hybrid and genotype decreased more the RMSE than adding only material type (Figure 4.6e, b). The *M. sin x M. sac* and *M. x giganteus* genotypes showed the highest improvement of RMSE due to the addition of these categorical variables (Figure 4.6f). The data imbalance in the "hybrid" categorical variables among control *M. x giganteus* (*n*=1), interspecies (*n*=4) and intraspecies (*n*=8) genotypes hybrids could have caused these differences in model's performance.

Another limitation of the RF model developed in this study relies on the fact that it is composed of multiple VIs calculated with precise multispectral bands. This means that our RF model might not reach the same accuracy if the same VIs are calculated on spectral data acquired with different multispectral cameras operating within different bands intervals. This calls for the development of algorithms able to overcome these differences in the spectral data through advanced normalization and calculation procedures of VIs from different sensors (Emilien *et al.*, 2021; Hoque & Phinn, 2018).

4.4.3 Phenotyping stay-green trait via UAV remote sensing to capture genotypic variation during senescence

This study demonstrated that high-throughput phenotyping (HTP) of contrasting *Miscanthus* hybrids is possible by combining multiple UAV flights and GAM modelling. Stay-green (SG) is an important phenotypic trait when evaluating the senescence of novel *Miscanthus* hybrids. The goal of plant breeders is to obtain high yielding plants with high biomass quality. In *Miscanthus*, a delayed senescence is expected to increase yields, while an early senescence is expected to increase biomass quality (Robson *et al.*, 2013). In our environments, senescence of *M. sin x M. sin* hybrids led to drier biomass (22% mean moisture content in late February) than commercially available rhizome-based hybrids (GRC 9-15 with 37%), while *M. sin x M. sac* hybrids showed a SG trait with an average moisture content of 45% until harvest. These findings confirmed that biomass with low moisture content at the harvest is usually related to early senescence in *Miscanthus*, as was found by Robson *et al.* (2013). However, opposite results to our study were reported by Nunn *et al.* (2017) that observed a lack of relationship between an early senescence and low moisture content at harvest in different locations across Europe.

Mild cold conditions during autumn-winter periods affected the start of senescence and moisture content losses dynamics until late winter harvest in all *Miscanthus* hybrids. The overwintering conditions (e.g. number and frequency of chilling frosts) between the start of senescence and harvest time have a higher effect on the moisture content than the senescence itself (Sarath *et al.*, 2014). That was the case in our two southern European locations, where a reduced frequency of killing frost days and absence of prolonged freezing periods in late autumn - early winter in 2019-2020 seasons (Table

4.1) might not have induced complete senescence in the *M. sin x M. sac* hybrids leading to a higher moisture content at harvest. During the first years after establishment, *Miscanthus* might have a reduced senescence (Clifton-Brown & Lewandowski, 2000) due to changes in the source-sink dynamics of young *Miscanthus* plants (Boersma *et al.*, 2015). However, standing age did not affect in our case the observed delayed senescence since measurements were done on mature plantation at 2nd and 3rd year.

Genotypic variations in flowering and senescence times are instead two key explanatory factors of SG trait observed in perennial crops. The relationship between flowering and senescence in *Miscanthus* has been proposed to promote nutrient remobilization, and hence biomass quality improvement (Jensen *et al.*, 2017). GAM applied to estimated moisture content values from the RF model from multiple UAV flights helped us to capture differences in senescence dynamics in contrasting *Miscanthus* hybrids (Figure 4.7). A constant increase in the differences between the estimated moisture content of interspecies *M. sin x M. sac* and intraspecies *M. sin x M. sin* hybrids was observed between DOY 300 (late October) and DOY 350 (mid-December). During this period, the mean temperature decreased under 10°C. Therefore, under these mean temperature conditions, *M. sin x M. sin* hybrids might be more sensitive to temperatures below 10°C and thus start active senescence sooner than *M. x giganteus* and *M. sin x M. sac* hybrids that instead showed a delayed senescence. Fonteyne *et al.* (2016) reviewed the effect of frost and chilling stress in *Miscanthus* genotypes and found that *M. sacchariflorus* was more resistant to cold stress than *M. sinensis.* All *M. sin x M. sac* genotype hybrids showed a more persistent SG compared to *M. x giganteus* and *M. sin x M. sin* hybrids. In agreement with our results, Rusinowski *et al.* (2019) found that *GNT 34* hybrid, (GRC 13 in this study), had a longer SG period than *M. x giganteus*. Only GRC 15 among *M. sin x M. sac* genotype had a similar senescence dynamic to the *M. x giganteus*. The similar senescence dynamics observed for these two commercially available rhizome-based hybrids confirms that transplanting material (rhizome *vs* seed-based plugs) has an impact on moisture content loss during senescence. The observed differences in senescence time and moisture content loss rate during senescence among *Miscanthus* genotypes are respectively linked to flowering time and nutrient remobilization. Other studies confirming that *M. sin x M. sin* hybrids flowered earlier (mid-summer) than rhizome-based *M. x giganteus* hybrid while *M. sin x M. sac* never reached flowering (Clifton-Brown & Lewandowski, 2002; Nunn *et al*., 2017). Jensen *et al*. (2017) found, for the similar contrasting hybrids, that nitrogen and phosphorous remobilization rate to underground rhizomes followed the same trend of moisture content loss observed also in our study. The absence or delay of flowering respectively in *M. sin x M. sac* and rhizome-based hybrids may have caused delayed senescence that was also observed in the SG trait in this study. As a consequence, these genotypes

were harvested at higher moisture content (Figure 4.2) and likely higher nutrient content compared to *M. sin x M. sin* hybrids. The high variability among *Miscanthus* hybrids in moisture content loss dynamics during senescence (Figure 4.7) might be further explained by the wider geographical distribution of *M. sinensis* than of *M. sacchariflorus* (Clifton-Brown *et al.*, 2015). This may have produced a higher genetic variation of the phenotypic traits due to the hybridisation among *M. sinensis* species (Robson *et al.*, 2013). Additionally, also the cold resistance trait likely depends on the origin and *in situ* environmental characteristics of the genetic accession of the *Miscanthus* species. In fact, opposite results to our study were reported by (Clifton-Brown *et al.*, 2002) showing that different *M. sin x M. sin* hybrids had delayed senescence with respect to *M. x giganteus* and *M. sacchariflorus* hybrids.

4.5 Conclusion

This study demonstrated that moisture content of *Miscanthus* can be accurately estimated via machine learning algorithm applied to multiple VIs calculated from UAV-based VIS-NIR images. The RF model developed on different genotypes showed a good transferability to multiple location and seasons when moisture content ranges from 30% to 60%. Further training datasets are required to extend the transferability and confirm the same performance of the RF model at lower moisture content values (10-30%). For the first time, we showed that the combination of machine learning (ML) and GAM applied to time series of moisture content values estimated from VIs derived from multiple UAV flights is a powerful tool for high throughput phenotyping (HTP). Remote sensing can be used for phenotyping future advanced breeding programs of *Miscanthus*. The possibility to distinguish via remote sensing the SG trait of novel *Miscanthus* hybrids can deepen our understanding of key factors mediating the induction of early or delayed senescence. Our study focused on the use ML algorithms to estimate moisture content during *Miscanthus* senescence, but we believe that the same methodological approach can be used for estimating other phenological traits or yield components in similar and/or different crops. This is particularly relevant for upscaling models from experimental plot to field scale by using satellites. Satellites can collect data of many fields simultaneously, with a larger number of spectral bands, like the SWIR band, that could ultimately support with high precision and resolution moisture content and yield estimation. ML algorithms could be applied in remote sensing to develop satellite and UAV applications beneficial to sustainable crop management e.g. in the case of *Miscanthus* to identify optimal harvest date or to predict commercial yield (quantity and quality).

Chapter 5

Synthesis

Synthesis

Abstract

This chapter first provides a synthesis of the main results presented in the previous chapters. Subsequently, in view of these results, the contributions to scientific context and future research are discussed.

Contents

5.1 Main results

The main objective of this thesis was to evaluate the application of the UAV multispectral remote sensing platform for high-throughput phenotyping of hemp and miscanthus traits. Seven research hypotheses were defined (section 1.5 Thesis outline and Table 1.2) and in this chapter each research hypothesis is addressed, reporting the main results.

H1. The ML models and PROSAIL model can be used to estimate traits of miscanthus hybrids and hemp cultivars.

This hypothesis was addressed in chapters 2, 3 and 4. The random forest (RF) algorithm was used in this thesis for estimation of miscanthus traits, such as moisture content (Chapter 4) and light interception, plant height, green leaf biomass and standing biomass (Chapter 3). This thesis demonstrated that the RF models can accurately estimate these crop traits using the VIs derived from UAV multispectral images, confirming the research hypothesis. Moisture content, light interception and plant height were estimated with better accuracy than green leaf biomass and standing biomass. The UAV multispectral remote sensing and RF models have proved to be suitable tools to estimate these crop traits for high-throughput phenotyping of novel miscanthus hybrids and can provide additional information for the calibration of perennial crop growth models, such as MISCANFOR. The research hypothesis was also confirmed for leaf area index (LAI) and leaf chlorophyll content (LCC) estimation of hemp using the inversion of the PROSAIL model (Chapter 2). The inversion of the PROSAIL model estimated the LCC trait with less accuracy than LAI, as reported in several studies. As ML models, the PROSAIL model can also estimate crop traits for high throughput phenotyping (HTP).

H2. The quality training data can develop robust ML models to overcome the transferability problem.

This hypothesis was addressed in chapter 4. To assess the RF models transferability, in this thesis the RF models trained with moisture content collected from a wide range of genotypes, across two senescence seasons and at two different locations differing strongly in soils and differing slightly in climate. To evaluate the RF models transferability, the moisture content dataset was split into 5 test datasets. The RF models showed a good transferability to multiple locations and seasons when moisture content ranges were from 30% to 60%, while it showed low transferability when the biomass of these genotypes dried to a moisture content <30%, confirming the transferability problem of ML models. Indeed, during the two seasons, many hybrids did not reach such a low moisture content, and thus the training set size for this interval was lower. However, the quality of the training data, in terms of the training set size and the moisture content distribution during senescence, was confirmed as

being the most important dataset characteristic to achieve good model transferability, confirming the research hypothesis.

H3. The peak derived from the VIs time series of UAV can be used to predict the yield of miscanthus.

This hypothesis was addressed in chapter 3. The random forest (RF) was trained with the peak derived from the time series of the VIs fitted via generalized additive model (GAM). The peak was able to predict the yield of miscanthus using the RF algorithm, proving to be an important land surface phenology (LSP) descriptor for yield prediction and confirming the research hypothesis. The peak for *Miscanthus* hybrids occurred on average in mid-summer in Italy (southern/warm) and early autumn in the UK (northern/cold) locations. The RF model accurately predicted the yield with 2.3 Mg DM ha^{-1} of RMSE and 19.7 % NRMSE. The operability of this model was evaluated by a timeline of the performance using the peak of VIs derived from partial time series. The RF model showed a good capability to predict the yield months before the harvest both in Italy and in the UK. This capability of the model is extremely relevant for optimising *Miscanthus* biomass supply chain logistics, from field to facilities creating bioproducts or biopower.

H4. The spectral bands used for calculating the VIs have different importance depending on the crop traits to be estimated.

This hypothesis was addressed in chapters 3 and 4. The variable importance of the RF models was calculated using a recursive feature elimination (RFE) algorithm (chapter 4) and by dropout loss of RMSE (chapter 3). This thesis found a link between the spectral bands used to calculate the VIs and the estimated or predicted crop traits, as suggested in several studies. These results confirmed the research hypothesis. Indeed, in chapter 3, the VIs based on the green band as greenWDRVI and GNDVI were the most important variables to estimate green leaf biomass and standing biomass and also to predict the miscanthus yield, as was found in switchgrass and other warm-season perennial grasses. In chapter 4, the most important VIs for moisture content estimation were blue (BNDVI), green (GDVI) and red-edge (PSRI) spectral bands-based VIs that were sensitive to changes of chlorophyll/carotenoid ratio during senescence. Indeed, the blue band is sensitive to changes in carotenoid content and the green and red-edge bands are sensitive to changes in chlorophyll content, as found in literature.

H5. The hybrid regression inversion methods will better estimate the crop traits than LUT inversion methods of the PROSAIL model.

This thesis compared different inversion methods: two look-up table (LUT) methods based on the $RMSE_r$ cost function (LUT-I and LUT-II) and four hybrid regression methods based on machine

learning techniques (RF, GPR, ANN and EM) for the estimation of LAI and LCC traits of two hemp cultivars, green and yellow, in two different seasons. Generally, all evaluated inversion methods showed good accuracies of the hemp traits, especially for LAI, proving a transferability in both seasons. The best accuracies were achieved by hybrid methods with RF for the LAI and GPR for the LCC trait. All the hybrid methods performed better than LUT methods for the LAI (except for the EM method) and LCC estimation, confirming the research hypothesis. Other studies compared hybrid and LUT methods, but there is no univocal opinion on the best inversion method.

H6. The PROSAIL model can be used to derive equations able to link multi-sensor VIs and to overcome the differences of VIs between sensors.

This hypothesis was addressed in chapter 3. The LUT used for the PROSAIL model inversion was generated from the miscanthus trait values found in literature and from the location characteristics (North-West Italy and Mid-West Wales). The equations derived from this LUT proved to be an interesting tool to overcome the issue of multi-sensor interoperability by linking VIs from multispectral sensors. This is particularly so when the objective is to estimate crop traits or to predict yield using VIs calculated from sensors with different spectral characteristics (MicaSense and SlantRange). The results showed that the importance of subjecting VIs to this linking procedure varies from VI to VI. In fact, the green and red-edge bands-based VIs require the use of linking equations because they showed higher variation than the red and NIR bands-based VIs. In this thesis, the MicaSense and the SlantRange sensors showed a difference in the green and red-edge bands. The MicaSense sensor red-edge bands had a different central wavelength and a narrower green FWHM (full width at half maximum) compared to the SlantRange sensor. Therefore, these linking procedures must only be used when spectral bands-based VIs with quite different characteristics are used.

H7. The GAM analysis, applied to the time series of the crop traits estimated by ML or PROSAIL model inversion, can be used for phenotyping the dynamics of the crop traits of contrasting miscanthus hybrids and hemp cultivars.

This hypothesis was addressed in chapters 2 and 4. This thesis demonstrated that the HTP obtained by combining multiple UAV flights, estimation models, such as the ML (chapter 4) and the inversion of the PROSAIL model (chapter 2) and GAM modelling can characterise the dynamics of the phenotypic crop traits, confirming the research hypothesis. The GAM applied to estimated hemp and miscanthus trait values by ML or inversion of the PROSAIL model from multiple UAV flights enabled contrasting hybrids or cultivars to be differentiated (Figure 4.7). In chapter 2, the GAM phenotyped the LAI and LCC dynamics between two hemp cultivars, yellow and green, under four nitrogen fertilisation levels and proved to be a useful tool to capture differences of LAI and LCC

dynamics throughout the whole growing season. In chapter 4, the GAM applied to time series of moisture content estimated-values phenotyped the stay-green trait and captured the differences in the senescence dynamics of miscanthus hybrids.

5.2 Scientific contribution

The present thesis contributes to improving the knowledge on adopting UAV multispectral remote sensing platform for HTP of biomass crops such as hemp and miscanthus and using this to support the field trials of breeding programs. The UAV multispectral remote sensing for HTP applications is a recent addition in agriculture, and it has already been performed on many crops. However, in this thesis, as well as evaluating the feasibility of using UAV-based HTP with hemp and miscanthus, it was also used as a tool to characterize the growth and senescence dynamics. Indeed, for the first time, it has been shown that the combination of estimation models (ML or inversion of the PROSAIL model) and GAM can phenotype the crop traits dynamics throughout the growing season, when applied to the time series of hemp and miscanthus trait values estimated from spectral data acquired from multiple UAV flights. It seems likely that this combination (UAV, ML or PROSAIL model, and GAM) will also be able to be used for the HTP of other crops in future breeding programs.

Of significant note is that in this thesis, for the first time the topic of UAV multi-sensor interoperability was explored; this had previously only been addressed for satellite platforms. This topic is very important as there are many UAV multispectral sensors on the market, each with different spectral characteristics. The VIs of two different sensors were calculated by resampling the reflectance of the canopy, simulated by the PROSAIL model, based on the different characteristics of the UAV sensors. The linking procedure used in this study is based on equations calculated through linear regression between the MicaSense VIs and the SlantRange VIs. These equations proved to be a powerful tool to overcome the issue of multi-sensor interoperability by linking VIs from the UAV multispectral sensors.

Finally, the peak derived from the VIs time series is widely used in satellite remote sensing for yield prediction. This thesis explored the potential use of this approach through UAV remote sensing to predict the yield of the novel miscanthus hybrids on small plots in field trials that cannot be monitored using satellite remote sensing. In addition, to evaluate the operability of the model, so its capability to predict the yield months before the harvest, a timeline of the performance of the model using peak derived from partial VIS time series was performed. The ability to early predict the yield of novel hybrids using UAV makes it a key ally for the establishment of a sustainable value chain based on biomass crops.

5.3 Future research

This thesis shows that UAV remote sensing is an important tool for HTP. The capacity to distinguish crop traits using UAV remote sensing can help to comprehend the differences between genotypes during the growing season. However, this thesis has also highlighted that there is a need for new future studies. Future UAV remote sensing studies should focus on evaluating the dynamics of other hemp traits such as stem biomass, leaf biomass or nitrogen content during the season. Unlike this thesis in which LAI and LCC were evaluated through the inversion of the PROSAIL model using UAV multispectral images, future studies could deal with the use of ML models, especially regarding traits such as stem, leaf biomass and nitrogen content for which the use of the inversion of the PROSAIL model would be unsuitable. Novel miscanthus studies could also be carried out in order to extend the RF model transferability for the moisture content estimation at values below 30%. The transferability of other machine learning algorithms should also be tested. To improve the miscanthus moisture content estimation, future studies could investigate the exploitation of the spectral bands of the satellites, such as the SWIR band, known to well predict crop moisture content. The methodological approach used in this thesis, which consists of the combination of UAV multiple multispectral flights, estimation models and GAM, could be used to estimate these or different traits in other crops. Further studies could use this approach to calibrate the input parameters of crop growth models or to explore the run-time calibration of models by integrating UAV remote sensing estimates of input parameters, as already done with satellite data.

HTP and yield prediction 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. Therefore, further studies based on satellite remote sensing are needed to predict yield of biomass crops like *Miscanthus* in order to evaluate their potential for commercial scale.

This thesis used a procedure based on equations derived from the PROSAIL model to link VIs calculated from UAV sensors with different spectral characteristics. Although this procedure does not consider the light and meteorological conditions during the flights which may affect the quality of the spectral data, it proved to be an interesting tool to overcome multi-sensor interoperability problems. Future studies could use this procedure to investigate the interoperability problems between the several UAV multispectral sensors available on the market. The use of this procedure could be

interesting for crop phenotyping, where the field trials are often carried out in multi-location and with different UAV sensors. Additionally, the multi-sensor interoperability topic could be relevant in future since the UAV scientific studies for agriculture applications are exponentially increasing, the UAV market is expanding, and the number of UAV users is constantly growing. Finally, this procedure could enable the building and the application of crop traits estimation models based on UAV multi-sensor data for developing remote sensing applications beneficial to sustainable crop management.

Reference

Adam, E., Mutanga, O., Abdel-Rahman, E. M., & Ismail, R. (2014). Estimating standing biomass in papyrus (Cyperus papyrus L.) swamp: exploratory of in situ hyperspectral indices and random forest regression. *International Journal of Remote Sensing*, *35*(2), 693–714. https://doi.org/10.1080/01431161.2013.870676

Agostini, A., Serra, P., Giuntoli, J., Martani, E., Ferrarini, A., & Amaducci, S. (2021). Biofuels from perennial energy crops on buffer strips: A win-win strategy. *Journal of Cleaner Production*, *297*, 126703. https://doi.org/10.1016/j.jclepro.2021.126703

Alam, M. M., Strandgard, M. N., Brown, M. W., & Fox, J. C. (2012). Improving the productivity of mechanised harvesting systems using remote sensing. *Australian Forestry*, *75*(4), 238–245. https://doi.org/10.1080/00049158.2012.10676408

Alexandridis, T. K., Ovakoglou, G., & Clevers, J. G. P. W. (2020). Relationship between MODIS EVI and LAI across time and space. *Geocarto International*, *35*(13), 1385–1399. https://doi.org/10.1080/10106049.2019.1573928

Ali, A. M., Darvishzadeh, R., Skidmore, A., Gara, T. W., & Heurich, M. (2021). Machine learning methods' performance in radiative transfer model inversion to retrieve plant traits from Sentinel-2 data of a mixed mountain forest. *International Journal of Digital Earth*, *14*(1), 106–120. https://doi.org/10.1080/17538947.2020.1794064

Amaducci, S., Scordia, D., Liu, F. H., Zhang, Q., Guo, H., Testa, G., & Cosentino, S. L. (2015). Key cultivation techniques for hemp in Europe and China. *Industrial Crops and Products*, *68*, 2–16. https://doi.org/10.1016/j.indcrop.2014.06.041

Amaducci, Stefano, Facciotto, G., Bergante, S., Perego, A., Serra, P., Ferrarini, A., & Chimento, C. (2017). Biomass production and energy balance of herbaceous and woody crops on marginal soils in the Po Valley. *GCB Bioenergy*, *9*(1), 31–45. https://doi.org/10.1111/gcbb.12341

Anderegg, J., Yu, K., Aasen, H., Walter, A., Liebisch, F., & Hund, A. (2020). Spectral Vegetation Indices to Track Senescence Dynamics in Diverse Wheat Germplasm. *Frontiers in Plant Science*, *10*, 1749. https://doi.org/10.3389/fpls.2019.01749

Antonucci, G., Croci, M., Miras-Moreno, B., Fracasso, A., & Amaducci, S. (2021). Integration of Gas Exchange With Metabolomics: High-Throughput Phenotyping Methods for Screening Biostimulant-Elicited Beneficial Responses to Short-Term Water Deficit. *Frontiers in Plant Science*, *12*. https://doi.org/10.3389/fpls.2021.678925

Araus, J. L., & Cairns, J. E. (2014). Field high-throughput phenotyping: the new crop breeding frontier. *Trends in Plant Science*, *19*(1), 52–61. https://doi.org/10.1016/j.tplants.2013.09.008

Asner, G. P. (1998). Biophysical and Biochemical Sources of Variability in Canopy Reflectance. *Remote Sensing of Environment*, *64*(3), 234–253. https://doi.org/10.1016/S0034-4257(98)00014-5

Atzberger, C. (2004). Object-based retrieval of biophysical canopy variables using artificial neural nets and radiative transfer models. *Remote Sensing of Environment*, *93*(1–2), 53–67. https://doi.org/10.1016/j.rse.2004.06.016

Atzberger, C. (2010). Inverting the PROSAIL canopy reflectance model using neural nets trained on streamlined databases. *Journal of Spectral Imaging*. https://doi.org/10.1255/jsi.2010.a2

Atzberger, C., Darvishzadeh, R., Immitzer, M., Schlerf, M., Skidmore, A., & le Maire, G. (2015). Comparative analysis of different retrieval methods for mapping grassland leaf area index using

airborne imaging spectroscopy. *International Journal of Applied Earth Observation and Geoinformation*, *43*, 19–31. https://doi.org/10.1016/j.jag.2015.01.009

Awty-Carroll, D., Al Hassan, M., Ashman, C., Magenau, E., Martani, E., Kontek, M., van der Pluijm, P., de Maupeou, E., McCalmont Chris Davey, J., van der Cruijsen, K., Jurišić, V., Amaducci, S., Lamy, I., Kam, J., Hoogendam, A., Dolstra, O., Ferrarini, A., Lewandowski, I., Trindade, L., … Clifton-Brown, J. (2022). Establishment and yield performance over three years of fourteen inter- and intra-species Miscanthus hybrids planted using novel agronomies at seven European sites. *GCB Bioenergy*, *submitted*.

Banerjee, B. P., Joshi, S., Thoday-Kennedy, E., Pasam, R. K., Tibbits, J., Hayden, M., Spangenberg, G., & Kant, S. (2020). High-throughput phenotyping using digital and hyperspectral imaging-derived biomarkers for genotypic nitrogen response. *Journal of Experimental Botany*, *71*(15), 4604–4615. https://doi.org/10.1093/jxb/eraa143

Bannari, A., Asalhi, H., & Teillet, P. M. (2002). Transformed difference vegetation index (TDVI) for vegetation cover mapping. *IEEE International Geoscience and Remote Sensing Symposium*, *5*, 3053–3055. https://doi.org/10.1109/IGARSS.2002.1026867

Baret, F., Jacquemoud, S., Guyot, G., & Leprieur, C. (1992). Modeled analysis of the biophysical nature of spectral shifts and comparison with information content of broad bands. *Remote Sensing of Environment*, *41*(2–3), 133–142. https://doi.org/10.1016/0034-4257(92)90073-S

Baret, Frédéric, & Buis, S. (2008). Estimating Canopy Characteristics from Remote Sensing Observations: Review of Methods and Associated Problems. In *Advances in Land Remote Sensing* (pp. 173–201). Springer Netherlands. https://doi.org/10.1007/978-1-4020-6450-0_7

Baret, FrÉderic, & Jacquemoud, S. (1994). Modeling Canopy Spectral Properties to Retrieve Biophysical and Biochemical Characteristics. In *Imaging Spectrometry — a Tool for Environmental Observations* (pp. 145–167). Springer Netherlands. https://doi.org/10.1007/978-0-585-33173-7_9

Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, *114*, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Berger, K., Atzberger, C., Danner, M., D'Urso, G., Mauser, W., Vuolo, F., Hank, T., D'Urso, G., Mauser, W., Vuolo, F., & Hank, T. (2018). Evaluation of the PROSAIL Model Capabilities for Future Hyperspectral Model Environments: A Review Study. *Remote Sensing*, *10*(2), 85. https://doi.org/10.3390/rs10010085

Berni, J., Zarco-Tejada, P. J., Suarez, L., & Fereres, E. (2009). Thermal and Narrowband Multispectral Remote Sensing for Vegetation Monitoring From an Unmanned Aerial Vehicle. *IEEE Transactions on Geoscience and Remote Sensing*, *47*(3), 722–738. https://doi.org/10.1109/TGRS.2008.2010457

Biecek, P. (2018). *DALEX: moDel Agnostic Language for Exploration and eXplanation*.

Birth, G. S., & McVey, G. R. (1968). Measuring the Color of Growing Turf with a Reflectance Spectrophotometer 1. *Agronomy Journal*, *60*(6), 640–643. https://doi.org/10.2134/agronj1968.00021962006000060016x

Blancon, J., Dutartre, D., Tixier, M.-H. H., Weiss, M., Comar, A., Praud, S., & Baret, F. (2019). A High-Throughput Model-Assisted Method for Phenotyping Maize Green Leaf Area Index Dynamics Using Unmanned Aerial Vehicle Imagery. *Frontiers in Plant Science*, *10*, 685. https://doi.org/10.3389/fpls.2019.00685

Blandinières, H. (2022). *Productivity, eco-physiology, and stem processability of a yellow hemp (Cannabis sativa L.) cultivar under varying levels of nitrogen fertilisation*.

Boe, A., & Beck, D. L. (2008). Yield Components of Biomass in Switchgrass. *Crop Science*, *48*(4), 1306–1311. https://doi.org/10.2135/cropsci2007.08.0482

Boersma, N. N., Dohleman, F. G., Miguez, F. E., & Heaton, E. A. (2015). Autumnal leaf senescence in Miscanthus × giganteus and leaf [N] differ by stand age. *Journal of Experimental Botany*, *66*(14), 4395–4401. https://doi.org/10.1093/jxb/erv129

Bogard, M., Jourdan, M., Allard, V., Martre, P., Perretant, M. R., Ravel, C., Heumez, E., Orford, S., Snape, J., Griffiths, S., Gaju, O., Foulkes, J., & Le Gouis, J. (2011). Anthesis date mainly explained correlations between post-anthesis leaf senescence, grain yield, and grain protein concentration in a winter wheat population segregating for flowering time QTLs. *Journal of Experimental Botany*, *62*(10), 3621–3636. https://doi.org/10.1093/jxb/err061

Broge, N. ., & Leblanc, E. (2001). Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. *Remote Sensing of Environment*, *76*(2), 156–172. https://doi.org/10.1016/S0034-4257(00)00197-8

Brown, M. E., Pinzon, J. E., Didan, K., Morisette, J. T., & Tucker, C. J. (2006). Evaluation of the consistency of long-term NDVI time series derived from AVHRR,SPOT-vegetation, SeaWiFS, MODIS, and Landsat ETM+ sensors. *IEEE Transactions on Geoscience and Remote Sensing*, *44*(7), 1787–1793. https://doi.org/10.1109/TGRS.2005.860205

Burczyk, H., Grabowska, L., Kołodziej, J., & Strybe, M. (2008). Industrial Hemp as a Raw Material for Energy Production. *Journal of Industrial Hemp*, *13*(1), 37–48. https://doi.org/10.1080/15377880801898717

Cabrera-Bosquet, L., Molero, G., Stellacci, A., Bort, J., Nogués, S., & Araus, J. (2011). NDVI as a potential tool for predicting biomass, plant nitrogen content and growth in wheat genotypes subjected to different water and nitrogen conditions. *Cereal Research Communications*, *39*(1), 147–159. https://doi.org/10.1556/CRC.39.2011.1.15

Carlsson, G., Mårtensson, L.-M., Prade, T., Svensson, S.-E., & Jensen, E. S. (2017). Perennial species mixtures for multifunctional production of biomass on marginal land. *GCB Bioenergy*, *9*(1), 191–201. https://doi.org/10.1111/gcbb.12373

Cerrudo, D., González Pérez, L., Mendoza Lugo, J., & Trachsel, S. (2017). Stay-Green and Associated Vegetative Indices to Breed Maize Adapted to Heat and Combined Heat-Drought Stresses. *Remote Sensing*, *9*(3), 235. https://doi.org/10.3390/rs9030235

Chen, J. M. (1996). Evaluation of Vegetation Indices and a Modified Simple Ratio for Boreal Applications. *Canadian Journal of Remote Sensing*, *22*(3), 229–242. https://doi.org/10.1080/07038992.1996.10855178

Christopher, J. T., Veyradier, M., Borrell, A. K., Harvey, G., Fletcher, S., & Chenu, K. (2014). Phenotyping novel stay-green traits to capture genetic variation in senescence dynamics. *Functional Plant Biology*, *41*(11), 1035. https://doi.org/10.1071/FP14052

Clifton-Brown, J. C., Breuer, J., & Jones, M. B. (2007). Carbon mitigation by the energy crop, Miscanthus. *Global Change Biology*, *13*(11), 2296–2307. https://doi.org/10.1111/j.1365-2486.2007.01438.x

Clifton-Brown, J. C. C., & Lewandowski, I. (2002). Screening Miscanthus genotypes in field trials to optimise biomass yield and quality in Southern Germany. *European Journal of Agronomy*, *16*(2), 97–110. https://doi.org/10.1016/S1161-0301(01)00120-4

Clifton-Brown, J. C., Lewandowski, I., Bangerth, F., & Jones, M. B. (2002). Comparative responses to water stress in stay-green, rapid- and slow senescing genotypes of the biomass crop, Miscanthus. *New Phytologist*, *154*(2), 335–345. https://doi.org/10.1046/J.1469-8137.2002.00381.X

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., Scurlock, J., Donnison, I. S., Glover, C., Gołąb, I., Greef, J. M., Gwyn, J., Harding, G., Hayes, C., Helios, W., … Flavell, R. (2017). Progress in upscaling *Miscanthus* biomass production for the European bio-economy with seed-based hybrids. *GCB Bioenergy*, *9*(1), 6–17. https://doi.org/10.1111/gcbb.12357

Clifton-Brown, J., Schwarz, K.-U., Awty-Carroll, D., Iurato, A., Meyer, H., Greef, J., Gwyn, J., Mos, M., Ashman, C., Hayes, C., Huang, L., Norris, J., Rodgers, C., Scordia, D., Shafiei, R., Squance, M., Swaller, T., Youell, S., Cosentino, S., … Robson, P. (2019). Breeding Strategies to Improve Miscanthus as a Sustainable Source of Biomass for Bioenergy and Biorenewable Products. *Agronomy*, *9*(11), 673. https://doi.org/10.3390/agronomy9110673

Clifton-Brown, Schwarz, & Hastings. (2015). History of the development of Miscanthus as a bioenergy crop: from small beginnings to potential realisation. *Biology and Environment: Proceedings of the Royal Irish Academy*, *115B*(1), 45. https://doi.org/10.3318/bioe.2015.05

Clifton-Brown, J. C., & Lewandowski, I. (2000). Overwintering problems of newly established Miscanthus plantations can be overcome by identifying genotypes with improved rhizome cold tolerance. *New Phytologist*, *148*(2), 287–294. https://doi.org/10.1046/j.1469-8137.2000.00764.x

Clifton-Brown, J. C., Lewandowski, I., Andersson, B., Basch, G., Christian, D. G., Kjeldsen, J. B., J∅rgensen, U., Mortensen, J. V., Riche, A. B., Schwarz, K.-U., Tayebi, K., & Teixeira, F. (2001). Performance of 15 Miscanthus Genotypes at Five Sites in Europe. *Agronomy Journal*, *93*(5), 1013–1019. https://doi.org/10.2134/agronj2001.9351013x

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., Bastien, C., Bopper, S., Botnari, V., Brancourt-Hulmel, M., Chen, Z., Clark, L. V., Cosentino, S., Dalton, S., Davey, C., … Lewandowski, I. (2019). Breeding progress and preparedness for mass-scale deployment of perennial lignocellulosic biomass crops switchgrass, miscanthus, willow and poplar. *GCB Bioenergy*, *11*(1), 118–151. https://doi.org/10.1111/gcbb.12566

Crini, G., & Lichtfouse, E. (Eds.). (2020). *Sustainable Agriculture Reviews 42* (Vol. 42). Springer International Publishing. https://doi.org/10.1007/978-3-030-41384-2

Crippen, R. (1990). Calculating the vegetation index faster. *Remote Sensing of Environment*, *34*(1), 71–73. https://doi.org/10.1016/0034-4257(90)90085-Z

Cui, Z., & Kerekes, J. P. (2018). Potential of Red Edge Spectral Bands in Future Landsat Satellites on Agroecosystem Canopy Green Leaf Area Index Retrieval. *Remote Sensing*, *10*(9), 1458. https://doi.org/10.3390/rs10091458

Darvishzadeh, R., Skidmore, A., Schlerf, M., & Atzberger, C. (2008). Inversion of a radiative transfer model for estimating vegetation LAI and chlorophyll in a heterogeneous grassland. *Remote Sensing of Environment*, *112*(5), 2592–2604. https://doi.org/10.1016/j.rse.2007.12.003

Datt, B. (1999). Remote Sensing of Water Content in Eucalyptus Leaves. *Australian Journal of Botany*, *47*(6), 909. https://doi.org/10.1071/BT98042

Dauber, J., Brown, C., Fernando, A. L., Finnan, J., Krasuska, E., Ponitka, J., Styles, D., Thrän, D., Van Groenigen, K. J., Weih, M., & Zah, R. (2012). Bioenergy from "surplus" land: environmental and socio-economic implications. *BioRisk*, *7*, 5–50.

https://doi.org/10.3897/biorisk.7.3036

Daughtry, C. S. T., Walthall, C. L., Kim, M. S., De Colstoun, E. B., & McMurtrey, J. E. (2000). Estimating Corn Leaf Chlorophyll Concentration from Leaf and Canopy Reflectance. *Remote Sensing of Environment*, *74*(2), 229–239. https://doi.org/10.1016/S0034-4257(00)00113-9

de Beurs, K. M., & Henebry, G. M. (2005). Land surface phenology and temperature variation in the International Geosphere-Biosphere Program high-latitude transects. *Global Change Biology*, *11*(5), 779–790. https://doi.org/10.1111/j.1365-2486.2005.00949.x

de Beurs, K. M., & Henebry, G. M. (2010). Spatio-Temporal Statistical Methods for Modelling Land Surface Phenology. In *Phenological Research* (pp. 177–208). Springer Netherlands. https://doi.org/10.1007/978-90-481-3335-2_9

de Castro, A. I., Shi, Y., Maja, J. M., & Peña, J. M. (2021). UAVs for Vegetation Monitoring: Overview and Recent Scientific Contributions. *Remote Sensing 2021, Vol. 13, Page 2139*, *13*(11), 2139. https://doi.org/10.3390/RS13112139

de Wit, A. J. W., & van Diepen, C. A. (2008). Crop growth modelling and crop yield forecasting using satellite-derived meteorological inputs. *International Journal of Applied Earth Observation and Geoinformation*, *10*(4), 414–425. https://doi.org/10.1016/j.jag.2007.10.004

Doktor, D., Lausch, A., Spengler, D., & Thurner, M. (2014). Extraction of Plant Physiological Status from Hyperspectral Signatures Using Machine Learning Methods. *Remote Sensing*, *6*(12), 12247–12274. https://doi.org/10.3390/rs61212247

Duan, B., Liu, Y., Gong, Y., Peng, Y., Wu, X., Zhu, R., & Fang, S. (2019). Remote estimation of rice LAI based on Fourier spectrum texture from UAV image. *Plant Methods*, *15*(1), 124. https://doi.org/10.1186/s13007-019-0507-8

Duan, S.-B. B., Li, Z.-L. L., Wu, H., Tang, B.-H. H., Ma, L., Zhao, E., & Li, C. (2014). Inversion of the PROSAIL model to estimate leaf area index of maize, potato, and sunflower fields from unmanned aerial vehicle hyperspectral data. *International Journal of Applied Earth Observation and Geoinformation*, *26*(1), 12–20. https://doi.org/10.1016/j.jag.2013.05.007

Eitel, J. U. H., Long, D. S., Gessler, P. E., & Smith, A. M. S. (2007). Using in-situ measurements to evaluate the new RapidEye ™ satellite series for prediction of wheat nitrogen status. *International Journal of Remote Sensing*, *28*(18), 4183–4190. https://doi.org/10.1080/01431160701422213

Emilien, A.-V., Thomas, C., & Thomas, H. (2021). UAV & satellite synergies for optical remote sensing applications: A literature review. *Science of Remote Sensing*, *3*, 100019. https://doi.org/10.1016/j.srs.2021.100019

Escadafal, R., & Huete, A. (1991). Improvement in remote sensing of low vegetation cover in arid regions by correcting vegetation indices for soil “noise.” *Comptes Rendus de l’Academie Des Sciences*.

Evans, F. H., & Shen, J. (2021). Long-Term Hindcasts of Wheat Yield in Fields Using Remotely Sensed Phenology, Climate Data and Machine Learning. *Remote Sensing 2021, Vol. 13, Page 2435*, *13*(13), 2435. https://doi.org/10.3390/RS13132435

Farrell, A. D., Clifton-Brown, J. C., Lewandowski, I., & Jones, M. B. (2006). Genotypic variation in cold tolerance influences the yield of Miscanthus. *Annals of Applied Biology*, *149*(3), 337–345. https://doi.org/10.1111/j.1744-7348.2006.00099.x

Fei, Y., Jiulin, S., Hongliang, F., Zuofang, Y., Jiahua, Z., Yunqiang, Z., Kaishan, S., Zongming, W., & Maogui, H. (2012). Comparison of different methods for corn LAI estimation over northeastern China. *International Journal of Applied Earth Observation and Geoinformation*,

18, 462–471. https://doi.org/10.1016/j.jag.2011.09.004

Feng, Lei, Chen, S., Zhang, C., Zhang, Y., & He, Y. (2021). A comprehensive review on recent applications of unmanned aerial vehicle remote sensing with various sensors for high-throughput plant phenotyping. *Computers and Electronics in Agriculture*, *182*, 106033. https://doi.org/10.1016/J.COMPAG.2021.106033

Feng, Luwei, Zhang, Z., Ma, Y., Du, Q., Williams, P., Drewry, J., & Luck, B. (2020). Alfalfa Yield Prediction Using UAV-Based Hyperspectral Imagery and Ensemble Learning. *Remote Sensing*, *12*(12), 2028. https://doi.org/10.3390/rs12122028

Ferchichi, A., Abbes, A. Ben, Barra, V., & Farah, I. R. (2022). Forecasting vegetation indices from spatio-temporal remotely sensed data using deep learning-based approaches: A systematic literature review. *Ecological Informatics*, 101552. https://doi.org/10.1016/j.ecoinf.2022.101552

Ferrarini, A., Martani, E., Fornasier, F., & Amaducci, S. (2021). High C input by perennial energy crops boosts belowground functioning and increases soil organic P content. *Agriculture, Ecosystems & Environment*, *308*, 107247. https://doi.org/10.1016/j.agee.2020.107247

Ferrarini, Andrea, Fornasier, F., Serra, P., Ferrari, F., Trevisan, M., & Amaducci, S. (2017). Impacts of willow and miscanthus bioenergy buffers on biogeochemical N removal processes along the soil-groundwater continuum. *GCB Bioenergy*, *9*(1), 246–261. https://doi.org/10.1111/gcbb.12340

Fonteyne, S., Roldán-Ruiz, I., Muylle, H., De Swaef, T., Reheul, D., & Lootens, P. (2016). A Review of Frost and Chilling Stress in Miscanthus and Its Importance to Biomass Yield. In *Perennial Biomass Crops for a Resource-Constrained World* (pp. 127–144). Springer International Publishing. https://doi.org/10.1007/978-3-319-44530-4_12

Fritsche, U. R., Sims, R. E. H., & Monti, A. (2010). Direct and indirect land-use competition issues for energy crops and their sustainable production - an overview. *Biofuels, Bioproducts and Biorefining*, *4*(6), 692–704. https://doi.org/10.1002/bbb.258

Furbank, R. T., & Tester, M. (2011). Phenomics – technologies to relieve the phenotyping bottleneck. *Trends in Plant Science*, *16*(12), 635–644. https://doi.org/10.1016/j.tplants.2011.09.005

Gallo, K., Ji, L., Reed, B., Eidenshink, J., & Dwyer, J. (2005). Multi-platform comparisons of MODIS and AVHRR normalized difference vegetation index data. *Remote Sensing of Environment*, *99*(3), 221–231. https://doi.org/10.1016/j.rse.2005.08.014

Gao, B. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, *58*(3), 257–266. https://doi.org/10.1016/S0034-4257(96)00067-3

Gelfand, I., Sahajpal, R., Zhang, X., Izaurralde, R. C., Gross, K. L., & Robertson, G. P. (2013). Sustainable bioenergy production from marginal lands in the US Midwest. *Nature*, *493*(7433), 514–517. https://doi.org/10.1038/nature11811

Gevaert, C. M., Suomalainen, J., Tang, J., & Kooistra, L. (2015). Generation of Spectral–Temporal Response Surfaces by Combining Multispectral Satellite and Hyperspectral UAV Imagery for Precision Agriculture Applications. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *8*(6), 3140–3146. https://doi.org/10.1109/JSTARS.2015.2406339

Gitelson, A. A. (2004). Wide Dynamic Range Vegetation Index for Remote Quantification of Biophysical Characteristics of Vegetation. *Journal of Plant Physiology*, *161*(2), 165–173. https://doi.org/10.1078/0176-1617-01176

Gitelson, A. A., Gritz †, Y., & Merzlyak, M. N. (2003). Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *Journal of Plant Physiology*, *160*(3), 271–282. https://doi.org/10.1078/0176-1617-00887

Gitelson, A. A., Kaufman, Y. J., & Merzlyak, M. N. (1996). Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sensing of Environment*, *58*(3), 289–298. https://doi.org/10.1016/S0034-4257(96)00072-7

Gitelson, A. A., Kaufman, Y. J., Stark, R., & Rundquist, D. (2002). Novel algorithms for remote estimation of vegetation fraction. *Remote Sensing of Environment*, *80*(1), 76–87. https://doi.org/10.1016/S0034-4257(01)00289-9

Gitelson, A. A., Keydan, G. P., & Merzlyak, M. N. (2006). Three-band model for noninvasive estimation of chlorophyll, carotenoids, and anthocyanin contents in higher plant leaves. *Geophysical Research Letters*, *33*(11), 2006GL026457. https://doi.org/10.1029/2006GL026457

Gitelson, A., & Merzlyak, M. N. (1994). Quantitative estimation of chlorophyll-a using reflectance spectra: Experiments with autumn chestnut and maple leaves. *Journal of Photochemistry and Photobiology B: Biology*, *22*(3), 247–252. https://doi.org/10.1016/1011-1344(93)06963-4

Gracia-Romero, A., Kefauver, S. C., Fernandez-Gallego, J. A., Vergara-Díaz, O., Nieto-Taladriz, M. T., & Araus, J. L. (2019). UAV and Ground Image-Based Phenotyping: A Proof of Concept with Durum Wheat. *Remote Sensing*, *11*(10), 1244. https://doi.org/10.3390/rs11101244

Guillen-Climent, M. L., Zarco-Tejada, P. J., & Villalobos, F. J. (2014). Estimating Radiation Interception in Heterogeneous Orchards Using High Spatial Resolution Airborne Imagery. *IEEE Geoscience and Remote Sensing Letters*, *11*(2), 579–583. https://doi.org/10.1109/LGRS.2013.2284660

Guo, W., Carroll, M. E., Singh, A., Swetnam, T. L., Merchant, N., Sarkar, S., Singh, A. K., & Ganapathysubramanian, B. (2021). UAS-Based Plant Phenotyping for Research and Breeding Applications. *Plant Phenomics*, *2021*, 1–21. https://doi.org/10.34133/2021/9840192

Guo, Y., Senthilnath, J., Wu, W., Zhang, X., Zeng, Z., & Huang, H. (2019). Radiometric Calibration for Multispectral Camera of Different Imaging Conditions Mounted on a UAV Platform. *Sustainability*, *11*(4), 978. https://doi.org/10.3390/su11040978

Haboudane, D, Miller, J. R., Pattey, E., Zarco-Tehada, P. J., & Strachan, I. B. (2004). Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sensing of Environment*, *90*(3), 337–352. https://doi.org/10.1016/j.rse.2003.12.013

Haboudane, Driss, Miller, J. R., Tremblay, N., Zarco-Tejada, P. J., & Dextraze, L. (2002). Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote Sensing of Environment*, *81*(2–3), 416–426. https://doi.org/10.1016/S0034-4257(02)00018-4

Hamada, Y., Zumpf, C. R., Cacho, J. F., Lee, D., Lin, C.-H., Boe, A., Heaton, E., Mitchell, R., Negri, M. C., Rescia, A., & Gmada, S. S. (2021). Remote Sensing-Based Estimation of Advanced Perennial Grass Biomass Yields for Bioenergy. *Land*, *10*(11), 1221. https://doi.org/10.3390/land10111221

Han-Ya, I., Ishii, K., & Noguchi, N. (2010). Satellite and Aerial Remote Sensing for Production Estimates and Crop Assessment. *Environment Control in Biology*, *48*(2), 51–58. https://doi.org/10.2525/ecb.48.51

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

Plant Methods, *15*(1), 10. https://doi.org/10.1186/s13007-019-0394-z

Hassan, M., Yang, M., Rasheed, A., Jin, X., Xia, X., Xiao, Y., & He, Z. (2018). Time-Series Multispectral Indices from Unmanned Aerial Vehicle Imagery Reveal Senescence Rate in Bread Wheat. *Remote Sensing*, *10*(6), 809. https://doi.org/10.3390/rs10060809

Hastings, A., Clifton-Brown, J., Wattenbach, M., Mitchell, C. P., & Smith, P. (2009). The development of MISCANFOR, a new Miscanthus crop growth model: towards more robust yield predictions under different climatic and soil conditions. *GCB Bioenergy*, *1*(2), 154–170. https://doi.org/10.1111/j.1757-1707.2009.01007.x

Hastings, A., Mos, M., Yesufu, J. A., McCalmont, J., Schwarz, K., Shafei, R., Ashman, C., Nunn, C., Schuele, H., Cosentino, S., Scalici, G., Scordia, D., Wagner, M., & Clifton-Brown, J. (2017). Economic and Environmental Assessment of Seed and Rhizome Propagated Miscanthus in the UK. *Frontiers in Plant Science*, *8*, 1058. https://doi.org/10.3389/fpls.2017.01058

Heaton, E. A., Dohleman, F. G., Miguez, A. F., Juvik, J. A., Lozovaya, V., Widholm, J., Zabotina, O. A., McIsaac, G. F., David, M. B., Voigt, T. B., Boersma, N. N., & Long, S. P. (2010). *Miscanthus* (pp. 75–137). https://doi.org/10.1016/B978-0-12-381518-7.00003-0

Henner, D. N., Hastings, A., Pogson, M., McNamara, N. P., Davies, C. A., & Smith, P. (2020). PopFor: A new model for estimating poplar yields. *Biomass and Bioenergy*, *134*, 105470. https://doi.org/10.1016/j.biombioe.2020.105470

Herppich, W. B., Gusovius, H.-J., Flemming, I., & Drastig, K. (2020). Effects of Drought and Heat on Photosynthetic Performance, Water Use and Yield of Two Selected Fiber Hemp Cultivars at a Poor-Soil Site in Brandenburg (Germany). *Agronomy*, *10*(9), 1361. https://doi.org/10.3390/agronomy10091361

Hodkinson, T. R., Chase, M. W., & Renvoize, S. A. (2002). Characterization of a Genetic Resource Collection for Miscanthus (Saccharinae, Andropogoneae, Poaceae) using AFLP and ISSR PCR. *Annals of Botany*, *89*(5), 627–636. https://doi.org/10.1093/aob/mcf091

Holloway, J., & Mengersen, K. (2018). Statistical Machine Learning Methods and Remote Sensing for Sustainable Development Goals: A Review. *Remote Sensing*, *10*(9), 1365. https://doi.org/10.3390/rs10091365

Hoque, M. A.-A., & Phinn, S. (2018). Methods for Linking Drone and Field Hyperspectral Data to Satellite Data. In *Fundamentals, Sensor Systems, Spectral Libraries, and Data Mining for Vegetation* (pp. 321–354). CRC Press. https://doi.org/10.1201/9781315164151-12

Huete, A. . (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, *25*(3), 295–309. https://doi.org/10.1016/0034-4257(88)90106-X

Huete, A., Didan, K., Miura, T., Rodriguez, E. ., Gao, X., & Ferreira, L. . (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, *83*(1–2), 195–213. https://doi.org/10.1016/S0034-4257(02)00096-2

Hunt, E. R., Daughtry, C. S. T., Eitel, J. U. H., & Long, D. S. (2011). Remote Sensing Leaf Chlorophyll Content Using a Visible Band Index. *Agronomy Journal*, *103*(4), 1090–1099. https://doi.org/10.2134/agronj2010.0395

Hunt, M. L., Blackburn, G. A., Carrasco, L., Redhead, J. W., & Rowland, C. S. (2019). High resolution wheat yield mapping using Sentinel-2. *Remote Sensing of Environment*, *233*, 111410. https://doi.org/10.1016/j.rse.2019.111410

Im, J., Jensen, J. R., Coleman, M., & Nelson, E. (2009). Hyperspectral remote sensing analysis of short rotation woody crops grown with controlled nutrient and irrigation treatments. *Geocarto*

International, *24*(4), 293–312. https://doi.org/10.1080/10106040802556207

Impollonia, G., Croci, M., Martani, E., Ferrarini, A., Kam, J., Trindade, L. M., Clifton-Brown, J., & Amaducci, S. (2022). Moisture content estimation and senescence phenotyping of novel Miscanthus hybrids combining UAV based remote sensing and machine learning. *GCB Bioenergy*, *submitted*.

Ivonyi, I., Zolton, I., & van der Werf, H. M. G. (1997). Influence of nitrogen supply and P and K levels of the soil on dry matter and nutrient accumulation of fiber hemp (Cannabis sativa L.). *Journal of the International Hemp Association*.

Jackson, T. J., Chen, D., Cosh, M., Li, F., Anderson, M., Walthall, C., Doriaswamy, P., & Hunt, E. R. (2004). Vegetation water content mapping using Landsat data derived normalized difference water index for corn and soybeans. *Remote Sensing of Environment*, *92*(4), 475–482. https://doi.org/10.1016/j.rse.2003.10.021

Jacquemoud, S., & Baret, F. (1990). PROSPECT: A model of leaf optical properties spectra. *Remote Sensing of Environment*, *34*(2), 75–91. https://doi.org/10.1016/0034-4257(90)90100-Z

Jacquemoud, Stéphane, Verhoef, W., Baret, F., Bacour, C., Zarco-Tejada, P. J., Asner, G. P., François, C., & Ustin, S. L. (2009). PROSPECT+SAIL models: A review of use for vegetation characterization. *Remote Sensing of Environment*, *113*, S56–S66. https://doi.org/10.1016/j.rse.2008.01.026

Jay, S., Maupas, F., Bendoula, R., & Gorretta, N. (2017). Retrieving LAI, chlorophyll and nitrogen contents in sugar beet crops from multi-angular optical remote sensing: Comparison of vegetation indices and PROSAIL inversion for field phenotyping. *Field Crops Research*, *210*, 33–46. https://doi.org/10.1016/j.fcr.2017.05.005

Jensen, E., Robson, P., Farrar, K., Thomas Jones, S., Clifton-Brown, J., Payne, R., Donnison, I., Jones, S. T., Clifton-Brown, J., Payne, R., & Donnison, I. (2017). Towards Miscanthus combustion quality improvement: the role of flowering and senescence. *GCB Bioenergy*, *9*(5), 891–908. https://doi.org/10.1111/gcbb.12391

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., & Kim, S.-H. (2016). Random Forests for Global and Regional Crop Yield Predictions. *PLOS ONE*, *11*(6), e0156571. https://doi.org/10.1371/journal.pone.0156571

Jeżowski, S. (2008). Yield traits of six clones of Miscanthus in the first 3 years following planting in Poland. *Industrial Crops and Products*, *27*(1), 65–68. https://doi.org/10.1016/j.indcrop.2007.07.013

Ji, Z., Pan, Y., Zhu, X., Wang, J., & Li, Q. (2021). Prediction of Crop Yield Using Phenological Information Extracted from Remote Sensing Vegetation Index. *Sensors*, *21*(4), 1406. https://doi.org/10.3390/s21041406

Jiang, Z., Huete, A., Didan, K., & Miura, T. (2008). Development of a two-band enhanced vegetation index without a blue band. *Remote Sensing of Environment*, *112*(10), 3833–3845. https://doi.org/10.1016/j.rse.2008.06.006

Jin, S., Su, Y., Gao, S., Hu, T., Liu, J., & Guo, Q. (2018). The Transferability of Random Forest in Canopy Height Estimation from Multi-Source Remote Sensing Data. *Remote Sensing*, *10*(8), 1183. https://doi.org/10.3390/rs10081183

Johansen, K., Morton, M. J. L., Malbeteau, Y., Aragon, B., Al-Mashharawi, S., Ziliani, M. G., Angel, Y., Fiene, G., Negrão, S., Mousa, M. A. A., Tester, M. A., & McCabe, M. F. (2020). Predicting Biomass and Yield in a Tomato Phenotyping Experiment Using UAV Imagery and Random Forest. *Frontiers in Artificial Intelligence*, *3*, 28. https://doi.org/10.3389/frai.2020.00028

Jones, M. B., Zimmermann, J., & Clifton-Brown, J. (2016). Long-Term Yields and Soil Carbon Sequestration from Miscanthus: A Review. In *Perennial Biomass Crops for a Resource-Constrained World* (pp. 43–49). Springer International Publishing. https://doi.org/10.1007/978-3-319-44530-4_4

Jones, Michael B., Finnan, J., & Hodkinson, T. R. (2015). Morphological and physiological traits for higher biomass production in perennial rhizomatous grasses grown on marginal land. *GCB Bioenergy*, *7*(2), 375–385. https://doi.org/10.1111/gcbb.12203

Jongschaap, R. E. E. (2006). Run-time calibration of simulation models by integrating remote sensing estimates of leaf area index and canopy nitrogen. *European Journal of Agronomy*, *24*(4), 316–324. https://doi.org/10.1016/j.eja.2005.10.009

Kamir, E., Waldner, F., & Hochman, Z. (2020). Estimating wheat yields in Australia using climate records, satellite image time series and machine learning methods. *ISPRS Journal of Photogrammetry and Remote Sensing*, *160*, 124–135. https://doi.org/10.1016/j.isprsjprs.2019.11.008

Kang, S., Post, W. M., Nichols, J. A., Wang, D., West, T. O., Bandaru, V., & Izaurralde, R. C. (2013). Marginal Lands: Concept, Assessment and Management. *Journal of Agricultural Science*, *5*(5). https://doi.org/10.5539/jas.v5n5p129

Kavats, O., Khramov, D., Sergieieva, K., & Vasyliev, V. (2019). Monitoring Harvesting by Time Series of Sentinel-1 SAR Data. *Remote Sensing*, *11*(21), 2496. https://doi.org/10.3390/rs11212496

Keating, B. ., Carberry, P. ., Hammer, G. ., Probert, M. ., Robertson, M. ., Holzworth, D., Huth, N. ., Hargreaves, J. N. ., Meinke, H., Hochman, Z., McLean, G., Verburg, K., Snow, V., Dimes, J. ., Silburn, M., Wang, E., Brown, S., Bristow, K. ., Asseng, S., … Smith, C. . (2003). An overview of APSIM, a model designed for farming systems simulation. *European Journal of Agronomy*, *18*(3–4), 267–288. https://doi.org/10.1016/S1161-0301(02)00108-9

Kim, Y., Huete, A. ., Miura, T., & Jiang, Z. (2010). Spectral compatibility of vegetation indices across sensors: band decomposition analysis with Hyperion data. *Journal of Applied Remote Sensing*, *4*(1), 043520. https://doi.org/10.1117/1.3400635

Kipp, S., Mistele, B., & Schmidhalter, U. (2014). Identification of stay-green and early senescence phenotypes in high-yielding winter wheat, and their relationship to grain yield and grain protein concentration using high-throughput phenotyping techniques. *Functional Plant Biology*, *41*(3), 227. https://doi.org/10.1071/FP13221

Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, *28*(5). https://doi.org/10.18637/jss.v028.i05

Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling*. Springer New York. https://doi.org/10.1007/978-1-4614-6849-3

Laliberte, A. S., Goforth, M. A., Steele, C. M., & Rango, A. (2011). Multispectral Remote Sensing from Unmanned Aircraft: Image Processing Workflows and Applications for Rangeland Environments. *Remote Sensing*, *3*(11), 2529–2551. https://doi.org/10.3390/rs3112529

Larsen, S. U., Jørgensen, U., Kjeldsen, J. B., & Lærke, P. E. (2014). Long-Term Miscanthus Yields Influenced by Location, Genotype, Row Distance, Fertilization and Harvest Season. *BioEnergy Research*, *7*(2), 620–635. https://doi.org/10.1007/s12155-013-9389-1

Lehnert, L. W., Meyer, H., Obermeier, W. A., Silva, B., Regeling, B., & Bendix, J. (2019). Hyperspectral Data Analysis in R : The hsdar Package. *Journal of Statistical Software*, *89*(12).

https://doi.org/10.18637/jss.v089.i12

Lewandowski, I., Clifton-Brown, J. C., Scurlock, J. M. O., & Huisman, W. (2000). Miscanthus: European experience with a novel energy crop. *Biomass and Bioenergy*, *19*(4), 209–227. https://doi.org/10.1016/S0961-9534(00)00032-5

Lewandowski, Iris, 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., Farrar, K., Fonteyne, S., Harding, G., Hastings, A., Huxley, L. M., Iqbal, Y., Khokhlov, N., Kiesel, A., Lootens, P., … Kalinina, O. (2016). Progress on Optimizing Miscanthus Biomass Production for the European Bioeconomy: Results of the EU FP7 Project OPTIMISC. *Frontiers in Plant Science*, *7*(NOVEMBER2016), 1620. https://doi.org/10.3389/fpls.2016.01620

Li, F., Piasecki, C., Millwood, R. J., Wolfe, B., Mazarei, M., & Stewart, C. N. (2020). High-Throughput Switchgrass Phenotyping and Biomass Modeling by UAV. *Frontiers in Plant Science*, *11*, 1532. https://doi.org/10.3389/fpls.2020.574073

Li, H., Yang, W., Lei, J., She, J., & Zhou, X. (2021). Estimation of leaf water content from hyperspectral data of different plant species by using three new spectral absorption indices. *PLOS ONE*, *16*(3), e0249351. https://doi.org/10.1371/journal.pone.0249351

Li, P., Jiang, L., & Feng, Z. (2013). Cross-Comparison of Vegetation Indices Derived from Landsat-7 Enhanced Thematic Mapper Plus (ETM+) and Landsat-8 Operational Land Imager (OLI) Sensors. *Remote Sensing*, *6*(1), 310–329. https://doi.org/10.3390/rs6010310

Liebisch, F., Kirchgessner, N., Schneider, D., Walter, A., & Hund, A. (2015). Remote, aerial phenotyping of maize traits with a mobile multi-sensor approach. *Plant Methods*, *11*(1), 9. https://doi.org/10.1186/s13007-015-0048-8

Liedtke, J. D., Hunt, C. H., George-Jaeggli, B., Laws, K., Watson, J., Potgieter, A. B., Cruickshank, A., & Jordan, D. R. (2020). High-Throughput Phenotyping of Dynamic Canopy Traits Associated with Stay-Green in Grain Sorghum. *Plant Phenomics*, *2020*, 1–10. https://doi.org/10.34133/2020/4635153

Liu, J., Shang, J., Qian, B., Huffman, T., Zhang, Y., Dong, T., Jing, Q., & Martin, T. (2019). Crop Yield Estimation Using Time-Series MODIS Data and the Effects of Cropland Masks in Ontario, Canada. *Remote Sensing*, *11*(20), 2419. https://doi.org/10.3390/rs11202419

Lopes, M. S., & Reynolds, M. P. (2012). Stay-green in spring wheat can be determined by spectral reflectance measurements (normalized difference vegetation index) independently from phenology. *Journal of Experimental Botany*, *63*(10), 3789–3798. https://doi.org/10.1093/jxb/ers071

Lv, Z., Meng, R., Man, J., Zeng, L., Wang, M., Xu, B., Gao, R., Sun, R., & Zhao, F. (2021). Modeling of winter wheat fAPAR by integrating Unmanned Aircraft Vehicle-based optical, structural and thermal measurement. *International Journal of Applied Earth Observation and Geoinformation*, *102*, 102407. https://doi.org/10.1016/j.jag.2021.102407

Machwitz, M., Pieruschka, R., Berger, K., Schlerf, M., Aasen, H., Fahrner, S., Jiménez-Berni, J., Baret, F., & Rascher, U. (2021). Bridging the Gap Between Remote Sensing and Plant Phenotyping—Challenges and Opportunities for the Next Generation of Sustainable Agriculture. *Frontiers in Plant Science*, *0*, 2334. https://doi.org/10.3389/FPLS.2021.749374

MacKerron, D. K. L., & Haverkort, A. J. (Eds.). (2004). *Decision support systems in potato production*. Wageningen Academic Publishers. https://doi.org/10.3920/978-90-8686-527-7

Magenau, E., Clifton-Brown, J., Awty-Carroll, D., Ashman, C., Ferrarini, A., Kontek, M., Martani, E., Roderick, K., Davey, C., Amaducci, S., Jurišić, V., Kam, J., Trindade, L., Lewandowski, I., & Kiesel, A. (2022). Site impacts nutrient translocation efficiency in intra- and interspecies

Miscanthus hybrids on marginal lands. *GCB Bioenergy*, *submitted*.

Makanza, R., Zaman-Allah, M., Cairns, J., Magorokosho, C., Tarekegne, A., Olsen, M., & Prasanna, B. (2018). High-Throughput Phenotyping of Canopy Cover and Senescence in Maize Field Trials Using Aerial Digital Canopy Imaging. *Remote Sensing*, *10*(2), 330. https://doi.org/10.3390/rs10020330

Malinowska, M., Donnison, I. S., & Robson, P. R. H. (2017). Phenomics analysis of drought responses in Miscanthus collected from different geographical locations. *GCB Bioenergy*, *9*(1), 78–91. https://doi.org/10.1111/gcbb.12350

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., Marcato Junior, J., Eduardo Teodoro, P., & Pistori, H. (2020). A random forest ranking approach to predict yield in maize with uav-based vegetation spectral indices. *Computers and Electronics in Agriculture*, *178*, 105791. https://doi.org/10.1016/j.compag.2020.105791

Martani, E., Ferrarini, A., Serra, P., Pilla, M., Marcone, A., & Amaducci, S. (2021). Belowground biomass C outweighs soil organic C of perennial energy crops: Insights from a long-term multispecies trial. *GCB Bioenergy*, *13*(3), 459–472. https://doi.org/10.1111/gcbb.12785

Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of machine-learning classification in remote sensing: an applied review. *International Journal of Remote Sensing*, *39*(9), 2784–2817. https://doi.org/10.1080/01431161.2018.1433343

Mayer, Z. (2019). *A Brief Introduction to caretEnsemble*.

McCalmont, J. P., Hastings, A., McNamara, N. P., Richter, G. M., Robson, P., Donnison, I. S., & Clifton-Brown, J. (2017). Environmental costs and benefits of growing Miscanthus for bioenergy in the UK. *GCB Bioenergy*, *9*(3), 489–507. https://doi.org/10.1111/gcbb.12294

Mehmood, M. A., Ibrahim, M., Rashid, U., Nawaz, M., Ali, S., Hussain, A., & Gull, M. (2017). Biomass production for bioenergy using marginal lands. *Sustainable Production and Consumption*, *9*, 3–21. https://doi.org/10.1016/j.spc.2016.08.003

Meijer, W. J. M., van der Werf, H. M. G., Mathijssen, E. W. J. M., & van den Brink, P. W. M. (1995). Constraints to dry matter production in fibre hemp (Cannabis sativa L.). *European Journal of Agronomy*, *4*(1), 109–117. https://doi.org/10.1016/S1161-0301(14)80022-1

Meroni, M., Atzberger, C., Vancutsem, C., Gobron, N., Baret, F., Lacaze, R., Eerens, H., & Leo, O. (2013). Evaluation of Agreement Between Space Remote Sensing SPOT-VEGETATION fAPAR Time Series. *IEEE Transactions on Geoscience and Remote Sensing*, *51*(4), 1951–1962. https://doi.org/10.1109/TGRS.2012.2212447

Meroni, M., Colombo, R., & Panigada, C. (2004). Inversion of a radiative transfer model with hyperspectral observations for LAI mapping in poplar plantations. *Remote Sensing of Environment*, *92*(2), 195–206. https://doi.org/10.1016/j.rse.2004.06.005

Meroni, Michele, d'Andrimont, R., Vrieling, A., Fasbender, D., Lemoine, G., Rembold, F., Seguini, L., & Verhegghen, A. (2021). Comparing land surface phenology of major European crops as derived from SAR and multispectral data of Sentinel-1 and -2. *Remote Sensing of Environment*, *253*, 112232. https://doi.org/10.1016/J.RSE.2020.112232

Meroni, Michele, Waldner, F., Seguini, L., Kerdiles, H., & Rembold, F. (2021). Yield forecasting with machine learning and small data: What gains for grains? *Agricultural and Forest Meteorology*, *308–309*, 108555. https://doi.org/10.1016/j.agrformet.2021.108555

Merzlyak, M. N., Gitelson, A. A., Chivkunova, O. B., & Rakitin, V. Y. (1999). Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. *Physiologia Plantarum*, *106*(1), 135–141. https://doi.org/10.1034/j.1399-3054.1999.106119.x

Millard, K., & Richardson, M. (2015). On the Importance of Training Data Sample Selection in Random Forest Image Classification: A Case Study in Peatland Ecosystem Mapping. *Remote Sensing*, *7*(7), 8489–8515. https://doi.org/10.3390/rs70708489

Miura, T., Yoshioka, H., Fujiwara, K., & Yamamoto, H. (2008). Inter-Comparison of ASTER and MODIS Surface Reflectance and Vegetation Index Products for Synergistic Applications to Natural Resource Monitoring. *Sensors*, *8*(4), 2480–2499. https://doi.org/10.3390/s8042480

Montazeaud, G., Karatoğma, H., Özturk, I., Roumet, P., Ecarnot, M., Crossa, J., Özer, E., Özdemir, F., & Lopes, M. S. (2016). Predicting wheat maturity and stay–green parameters by modeling spectral reflectance measurements and their contribution to grain yield under rainfed conditions. *Field Crops Research*, *196*, 191–198. https://doi.org/10.1016/j.fcr.2016.06.021

Munaiz, E. D., Martínez, S., Kumar, A., Caicedo, M., & Ordás, B. (2020). The Senescence (Stay-Green)—An Important Trait to Exploit Crop Residuals for Bioenergy. *Energies*, *13*(4), 790. https://doi.org/10.3390/en13040790

Nolè, A., Rita, A., Ferrara, A. M. S., & Borghetti, M. (2018). 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. *Annals of Forest Science*, *75*(3), 83. https://doi.org/10.1007/s13595-018-0763-1

Nunn, C., Hastings, A. F. S. J., Kalinina, O., Özgüven, M., Schüle, H., Tarakanov, I. G., Van Der Weijde, T., Anisimov, A. A., Iqbal, Y., Kiesel, A., Khokhlov, N. F., McCalmont, J. P., Meyer, H., Mos, M., Schwarz, K.-U., Trindade, L. M., Lewandowski, I., & Clifton-Brown, J. C. (2017). Environmental Influences on the Growing Season Duration and Ripening of Diverse Miscanthus Germplasm Grown in Six Countries. *Frontiers in Plant Science*, *8*, 907. https://doi.org/10.3389/fpls.2017.00907

Ostos-Garrido, F. J., de Castro, A. I., Torres-Sánchez, J., Pistón, F., & Peña, J. M. (2019). High-Throughput Phenotyping of Bioethanol Potential in Cereals Using UAV-Based Multi-Spectral Imagery. *Frontiers in Plant Science*, *10*, 948. https://doi.org/10.3389/fpls.2019.00948

Pancaldi, F., & Trindade, L. M. (2020). Marginal Lands to Grow Novel Bio-Based Crops: A Plant Breeding Perspective. *Frontiers in Plant Science*, *11*. https://doi.org/10.3389/fpls.2020.00227

Parmley, K. A., Higgins, R. H., Ganapathysubramanian, B., Sarkar, S., & Singh, A. K. (2019). Machine Learning Approach for Prescriptive Plant Breeding. *Scientific Reports*, *9*(1), 17132. https://doi.org/10.1038/s41598-019-53451-4

Peng, B., Guan, K., Pan, M., & Li, Y. (2018). Benefits of Seasonal Climate Prediction and Satellite Data for Forecasting U.S. Maize Yield. *Geophysical Research Letters*, *45*(18), 9662–9671. https://doi.org/10.1029/2018GL079291

Penuelas, J., Baret, F., & Filella, I. (1995). Semi-empirical indices to assess carotenoids/chlorophyll a ratio from leaf spectral reflectance. *Photosynthetica*.

Peñuelas, J., & Inoue, Y. (1999). Reflectance Indices Indicative of Changes in Water and Pigment Contents of Peanut and Wheat Leaves. *Photosynthetica*, *36*(3), 355–360. https://doi.org/10.1023/A:1007033503276

Piotrowski, S., Carus, M., & Essel, R. (2015). Global Bioeconomy in the Conflict Between Biomass Supply and Demand. *Industrial Biotechnology*, *11*(6), 308–315. https://doi.org/10.1089/ind.2015.29021.stp

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. (2017). Multi-Spectral Imaging from an Unmanned Aerial Vehicle Enables the Assessment of Seasonal Leaf Area Dynamics of Sorghum Breeding Lines. *Frontiers in Plant Science*, *8*, 1532. https://doi.org/10.3389/fpls.2017.01532

Prasad, N. R., Patel, N. R., & Danodia, A. (2021). Cotton Yield Estimation Using Phenological Metrics Derived from Long-Term MODIS Data. *Journal of the Indian Society of Remote Sensing*, *49*(11), 2597–2610. https://doi.org/10.1007/s12524-021-01414-6

Prévot, L., Chauki, H., Troufleau, D., Weiss, M., Baret, F., & Brisson, N. (2003). Assimilating optical and radar data into the STICS crop model for wheat. *Agronomie*, *23*(4), 297–303. https://doi.org/10.1051/agro:2003003

Psomiadis, E., Dercas, N., Dalezios, N. R., & Spiropoulos, N. V. (2017). Evaluation and cross-comparison of vegetation indices for crop monitoring from sentinel-2 and worldview-2 images. In C. M. Neale & A. Maltese (Eds.), *Remote Sensing for Agriculture, Ecosystems, and Hydrology XIX* (p. 79). SPIE. https://doi.org/10.1117/12.2278217

Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H., & Sorooshian, S. (1994). A modified soil adjusted vegetation index. *Remote Sensing of Environment*, *48*(2), 119–126. https://doi.org/10.1016/0034-4257(94)90134-1

Quinn, L. D., Straker, K. C., Guo, J., Kim, S., Thapa, S., Kling, G., Lee, D. K., & Voigt, T. B. (2015). Stress-Tolerant Feedstocks for Sustainable Bioenergy Production on Marginal Land. *BioEnergy Research*, *8*(3), 1081–1100. https://doi.org/10.1007/s12155-014-9557-y

Rahaman, M. M., Chen, D., Gillani, Z., Klukas, C., & Chen, M. (2015). Advanced phenotyping and phenotype data analysis for the study of plant growth and development. *Frontiers in Plant Science*, *6*. https://doi.org/10.3389/fpls.2015.00619

Rengarajan, R., & Schott, J. (2018). Evaluation of Sensor and Environmental Factors Impacting the Use of Multiple Sensor Data for Time-Series Applications. *Remote Sensing*, *10*(11), 1678. https://doi.org/10.3390/rs10111678

Richardson, A. J., & Wiegand, C. L. (1977). Distinguishing Vegetation from Soil Background Information. *Photogrammetric Engineering and Remote Sensing*, *43*, 1541–1552.

Richter, G. M., Agostini, F., Barker, A., Costomiris, D., & Qi, A. (2016). Assessing on-farm productivity of Miscanthus crops by combining soil mapping, yield modelling and remote sensing. *Biomass and Bioenergy*, *85*, 252–261. https://doi.org/10.1016/j.biombioe.2015.12.024

Ritchie, R. J. (2006). Consistent Sets of Spectrophotometric Chlorophyll Equations for Acetone, Methanol and Ethanol Solvents. *Photosynthesis Research*, *89*(1), 27–41. https://doi.org/10.1007/s11120-006-9065-9

Robson, P. R. H., Farrar, K., Gay, A. P., Jensen, E. F., Clifton-Brown, J. C., & Donnison, I. S. (2013). Variation in canopy duration in the perennial biofuel crop Miscanthus reveals complex associations with yield. *Journal of Experimental Botany*, *64*(8), 2373–2383. https://doi.org/10.1093/jxb/ert104

Rondeaux, G., Steven, M., & Baret, F. (1996). Optimization of soil-adjusted vegetation indices. *Remote Sensing of Environment*, *55*(2), 95–107. https://doi.org/10.1016/0034-4257(95)00186-7

Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1973). *Monitoring vegetation systems in the great plains with ERTS*.

Rusinowski, S., Krzyżak, J., Clifton-Brown, J., Jensen, E., Mos, M., Webster, R., Sitko, K., &

Pogrzeba, M. (2019). New Miscanthus hybrids cultivated at a Polish metal-contaminated site demonstrate high stomatal regulation and reduced shoot Pb and Cd concentrations. *Environmental Pollution*, *252*, 1377–1387. https://doi.org/10.1016/j.envpol.2019.06.062

Sakamoto, T., Gitelson, A. A., & Arkebauer, T. J. (2013). MODIS-based corn grain yield estimation model incorporating crop phenology information. *Remote Sensing of Environment*, *131*, 215–231. https://doi.org/10.1016/j.rse.2012.12.017

Salas Fernandez, M. G., Becraft, P. W., Yin, Y., & Lübberstedt, T. (2009). From dwarves to giants? Plant height manipulation for biomass yield. *Trends in Plant Science*, *14*(8), 454–461. https://doi.org/10.1016/j.tplants.2009.06.005

Samuelsson, R., Burvall, J., & Jirjis, R. (2006). Comparison of different methods for the determination of moisture content in biomass. *Biomass and Bioenergy*, *30*(11), 929–934. https://doi.org/10.1016/j.biombioe.2006.06.004

Sarath, G., Baird, L. M., & Mitchell, R. B. (2014). Senescence, dormancy and tillering in perennial C4 grasses. *Plant Science*, *217–218*, 140–151. https://doi.org/10.1016/j.plantsci.2013.12.012

Schauberger, B., Jägermeyr, J., & Gornott, C. (2020). A systematic review of local to regional yield forecasting approaches and frequently used data resources. *European Journal of Agronomy*, *120*, 126153. https://doi.org/10.1016/j.eja.2020.126153

Schmidt, T., Fernando, A. L., Monti, A., & Rettenmaier, N. (2015). Life Cycle Assessment of Bioenergy and Bio-Based Products from Perennial Grasses Cultivated on Marginal Land in the Mediterranean Region. *BioEnergy Research*, *8*(4), 1548–1561. https://doi.org/10.1007/s12155-015-9691-1

Segarra, J., Buchaillot, M. L., Araus, J. L., & Kefauver, S. C. (2020). Remote Sensing for Precision Agriculture: Sentinel-2 Improved Features and Applications. *Agronomy*, *10*(5), 641. https://doi.org/10.3390/agronomy10050641

Sehgal, V. K., Chakraborty, D., & Sahoo, R. N. (2016). Inversion of radiative transfer model for retrieval of wheat biophysical parameters from broadband reflectance measurements. *Information Processing in Agriculture*, *3*(2), 107–118. https://doi.org/10.1016/j.inpa.2016.04.001

Seleiman, M. F., Santanen, A., Jaakkola, S., Ekholm, P., Hartikainen, H., Stoddard, F. L., & Mäkelä, P. S. A. (2013). Biomass yield and quality of bioenergy crops grown with synthetic and organic fertilizers. *Biomass and Bioenergy*, *59*, 477–485. https://doi.org/10.1016/j.biombioe.2013.07.021

Senthilnath, J., Dokania, A., Kandukuri, M., K.N., R., Anand, G., & Omkar, S. N. N. (2016). Detection of tomatoes using spectral-spatial methods in remotely sensed RGB images captured by UAV. *Biosystems Engineering*, *146*, 16–32. https://doi.org/10.1016/j.biosystemseng.2015.12.003

Shah, S. H., Angel, Y., Houborg, R., Ali, S., & McCabe, M. F. (2019). A Random Forest Machine Learning Approach for the Retrieval of Leaf Chlorophyll Content in Wheat. *Remote Sensing*, *11*(8), 920. https://doi.org/10.3390/rs11080920

She, X., Zhang, L., Cen, Y., Wu, T., Huang, C., & Baig, M. H. A. (2015). Comparison of the Continuity of Vegetation Indices Derived from Landsat 8 OLI and Landsat 7 ETM+ Data among Different Vegetation Types. *Remote Sensing*, *7*(10), 13485–13506. https://doi.org/10.3390/rs71013485

Shepherd, A., Clifton-Brown, J., Kam, J., Buckby, S., & Hastings, A. (2020). Commercial experience with miscanthus crops: Establishment, yields and environmental observations. *GCB Bioenergy*, *12*(7), 510–523. https://doi.org/10.1111/gcbb.12690

Shi, Y., Thomasson, J. A., Murray, S. C., Pugh, N. A., Rooney, W. L., Shafian, S., Rajan, N., Rouze, G., Morgan, C. L. S., Neely, H. L., Rana, A., Bagavathiannan, M. V., Henrickson, J., Bowden, E., Valasek, J., Olsenholler, J., Bishop, M. P., Sheridan, R., Putman, E. B., … Yang, C. (2016). Unmanned Aerial Vehicles for High-Throughput Phenotyping and Agronomic Research. *PLOS ONE*, *11*(7), e0159781. https://doi.org/10.1371/journal.pone.0159781

Shortall, O. K. (2013). "Marginal land" for energy crops: Exploring definitions and embedded assumptions. *Energy Policy*, *62*, 19–27. https://doi.org/10.1016/j.enpol.2013.07.048

Sishodia, R. P., Ray, R. L., & Singh, S. K. (2020). Applications of Remote Sensing in Precision Agriculture: A Review. *Remote Sensing*, *12*(19), 3136. https://doi.org/10.3390/rs12193136

Smith, R., Adams, J., Stephens, D., & Hick, P. (1995). Forecasting wheat yield in a Mediterranean-type environment from the NOAA satellite. *Australian Journal of Agricultural Research*, *46*(1), 113. https://doi.org/10.1071/AR9950113

Sripada, R. P., Heiniger, R. W., White, J. G., & Meijer, A. D. (2006). Aerial Color Infrared Photography for Determining Early In-Season Nitrogen Requirements in Corn. *Agronomy Journal*, *98*(4), 968–977. https://doi.org/10.2134/agronj2005.0200

Struik, P. C., Amaducci, S., Bullard, M. J., Stutterheim, N. C., Venturi, G., & Cromack, H. T. H. (2000). Agronomy of fibre hemp (Cannabis sativa L.) in Europe. *Industrial Crops and Products*, *11*(2–3), 107–118. https://doi.org/10.1016/S0926-6690(99)00048-5

Styks, J., Wróbel, M., Frączek, J., & Knapczyk, A. (2020). Effect of Compaction Pressure and Moisture Content on Quality Parameters of Perennial Biomass Pellets. *Energies*, *13*(8), 1859. https://doi.org/10.3390/en13081859

Su, W., Zhang, M., Bian, D., Liu, Z., Huang, J., Wang, W., Wu, J., & Guo, H. (2019). Phenotyping of Corn Plants Using Unmanned Aerial Vehicle (UAV) Images. *Remote Sensing*, *11*(17), 2021. https://doi.org/10.3390/rs11172021

Sun, B., Wang, C., Yang, C., Xu, B., Zhou, G., Li, X., Xie, J., Xu, S., Liu, B., Xie, T., Kuai, J., & Zhang, J. (2021). Retrieval of rapeseed leaf area index using the PROSAIL model with canopy coverage derived from UAV images as a correction parameter. *International Journal of Applied Earth Observation and Geoinformation*, *102*, 102373. https://doi.org/10.1016/j.jag.2021.102373

Tang, K., Struik, P. C., Yin, X., Calzolari, D., Musio, S., Thouminot, C., Bjelková, M., Stramkale, V., Magagnini, G., & Amaducci, S. (2017). A comprehensive study of planting density and nitrogen fertilization effect on dual-purpose hemp (Cannabis sativa L.) cultivation. *Industrial Crops and Products*, *107*, 427–438. https://doi.org/10.1016/j.indcrop.2017.06.033

Tang, K., Struik, P. C., Yin, X., Thouminot, C., Bjelková, M., Stramkale, V., & Amaducci, S. (2016). Comparing hemp (Cannabis sativa L.) cultivars for dual-purpose production under contrasting environments. *Industrial Crops and Products*, *87*, 33–44. https://doi.org/10.1016/j.indcrop.2016.04.026

Tao, H., Feng, H., Xu, L., Miao, M., Yang, G., Yang, X., & Fan, L. (2020). Estimation of the Yield and Plant Height of Winter Wheat Using UAV-Based Hyperspectral Images. *Sensors*, *20*(4), 1231. https://doi.org/10.3390/s20041231

Teillet, P., Fedosejevs, G., Barker, J., Miskey, C., & Bannari, A. (2006). Spectral Simulations of Vegetation Indices in the Context of Landsat Data Continuity. *2006 IEEE International Symposium on Geoscience and Remote Sensing*, 1784–1787. https://doi.org/10.1109/IGARSS.2006.461

Teillet, P. M., & Ren, X. (2008). Spectral band difference effects on vegetation indices derived from multiple satellite sensor data. *Canadian Journal of Remote Sensing*, *34*, 159–173.

Tejera, M. D., Miguez, F. E., & Heaton, E. A. (2021). The older plant gets the sun: Age-related changes in Miscanthus × giganteus phenology. *GCB Bioenergy*, *13*(1), 4–20. https://doi.org/10.1111/GCBB.12745

Tester, M., & Langridge, P. (2010). Breeding Technologies to Increase Crop Production in a Changing World. *Science*, *327*(5967), 818–822. https://doi.org/10.1126/science.1183700

Théau, J., Sankey, T. T., & Weber, K. T. (2010). Multi-sensor Analyses of Vegetation Indices in a Semi-arid Environment. *GIScience & Remote Sensing*, *47*(2), 260–275. https://doi.org/10.2747/1548-1603.47.2.260

Thomas, H., & Howarth, C. J. (2000). Five ways to stay green. *Journal of Experimental Botany*, *51*(suppl_1), 329–337. https://doi.org/10.1093/jexbot/51.suppl_1.329

Thouminot, C. (2015). La sélection française du chanvre : panorama et perspectives. *OCL*, *22*(6), D603. https://doi.org/10.1051/ocl/2015044

Tillack, A., Clasen, A., Kleinschmit, B., & Förster, M. (2014). Estimation of the seasonal leaf area index in an alluvial forest using high-resolution satellite-based vegetation indices. *Remote Sensing of Environment*, *141*, 52–63. https://doi.org/10.1016/j.rse.2013.10.018

Tilman, D., Socolow, R., Foley, J. A., Hill, J., Larson, E., Lynd, L., Pacala, S., Reilly, J., Searchinger, T., Somerville, C., & Williams, R. (2009). Beneficial Biofuels—The Food, Energy, and Environment Trilemma. *Science*, *325*(5938), 270–271. https://doi.org/10.1126/science.1177970

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, *8*(2), 127–150. https://doi.org/10.1016/0034-4257(79)90013-0

Upreti, D., Huang, W., Kong, W., Pascucci, S., Pignatti, S., Zhou, X., Ye, H., & Casa, R. (2019). A Comparison of Hybrid Machine Learning Algorithms for the Retrieval of Wheat Biophysical Variables from Sentinel-2. *Remote Sensing*, *11*(5), 481. https://doi.org/10.3390/rs11050481

Urrego, J. P. F., Huang, B., Næss, J. S., Hu, X., & Cherubini, F. (2021). Meta-analysis of leaf area index, canopy height and root depth of three bioenergy crops and their effects on land surface modeling. *Agricultural and Forest Meteorology*, *306*, 108444. https://doi.org/10.1016/j.agrformet.2021.108444

van der Cruijsen, K., Al Hassan, M., van Erven, G., Dolstra, O., & Trindade, L. M. (2021). Breeding Targets to Improve Biomass Quality in Miscanthus. *Molecules*, *26*(2), 254. https://doi.org/10.3390/molecules26020254

van der Werf, H. M. G. (2004). Life Cycle Analysis of field production of fibre hemp, the effect of production practices on environmental impacts. *Euphytica*, *140*(1–2), 13–23. https://doi.org/10.1007/s10681-004-4750-2

van Leeuwen, W. J. D., Orr, B. J., Marsh, S. E., & Herrmann, S. M. (2006). Multi-sensor NDVI data continuity: Uncertainties and implications for vegetation monitoring applications. *Remote Sensing of Environment*, *100*(1), 67–81. https://doi.org/10.1016/j.rse.2005.10.002

Varela, S., Pederson, T., Bernacchi, C. J., & Leakey, A. D. B. (2021). Understanding Growth Dynamics and Yield Prediction of Sorghum Using High Temporal Resolution UAV Imagery Time Series and Machine Learning. *Remote Sensing*, *13*(9), 1763. https://doi.org/10.3390/rs13091763

Venturi, P., Amaducci, S., Amaducci, M. T., & Venturi, G. (2007). Interaction Between Agronomic and Mechanical Factors for Fiber Crops Harvesting: Italian Results-Note II. Hemp. *Journal of Natural Fibers*, *4*(3), 83–97. https://doi.org/10.1300/J395v04n03_06

Verhoef, W. (1984). Light scattering by leaf layers with application to canopy reflectance modeling: The SAIL model. *Remote Sensing of Environment*, *16*(2), 125–141. https://doi.org/10.1016/0034-4257(84)90057-9

Verrelst, J., Camps-Valls, G., Muñoz-Marí, J., Rivera, J. P., Veroustraete, F., Clevers, J. G. P. W., & Moreno, J. (2015). Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties – A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *108*, 273–290. https://doi.org/10.1016/j.isprsjprs.2015.05.005

Verrelst, J., Dethier, S., Rivera, J. P., Munoz-Mari, J., Camps-Valls, G., & Moreno, J. (2016). Active Learning Methods for Efficient Hybrid Biophysical Variable Retrieval. *IEEE Geoscience and Remote Sensing Letters*, *13*(7), 1012–1016. https://doi.org/10.1109/LGRS.2016.2560799

Verrelst, J., Malenovský, Z., Van der Tol, C., Camps-Valls, G., Gastellu-Etchegorry, J.-P., Lewis, P., North, P., & Moreno, J. (2019). Quantifying Vegetation Biophysical Variables from Imaging Spectroscopy Data: A Review on Retrieval Methods. *Surveys in Geophysics*, *40*(3), 589–629. https://doi.org/10.1007/s10712-018-9478-y

Verrelst, J., Muñoz, J., Alonso, L., Delegido, J., Rivera, J. P., Camps-Valls, G., & Moreno, J. (2012). Machine learning regression algorithms for biophysical parameter retrieval: Opportunities for Sentinel-2 and -3. *Remote Sensing of Environment*, *118*, 127–139. https://doi.org/10.1016/j.rse.2011.11.002

Verrelst, J., Rivera, J. P., Leonenko, G., Alonso, L., & Moreno, J. J. (2014). Optimizing LUT-Based RTM Inversion for Semiautomatic Mapping of Crop Biophysical Parameters from Sentinel-2 and -3 Data: Role of Cost Functions. *IEEE Transactions on Geoscience and Remote Sensing*, *52*(1), 257–269. https://doi.org/10.1109/TGRS.2013.2238242

Verrelst, J., Rivera, J. P., Veroustraete, F., Muñoz-Marí, J., Clevers, J. G. P. W., Camps-Valls, G., & Moreno, J. (2015). Experimental Sentinel-2 LAI estimation using parametric, non-parametric and physical retrieval methods – A comparison. *ISPRS Journal of Photogrammetry and Remote Sensing*, *108*, 260–272. https://doi.org/10.1016/j.isprsjprs.2015.04.013

Villaescusa-Nadal, J. L., Franch, B., Roger, J.-C., Vermote, E. F., Skakun, S., & Justice, C. (2019). Spectral Adjustment Model's Analysis and Application to Remote Sensing Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *12*(3), 961–972. https://doi.org/10.1109/JSTARS.2018.2890068

Vincini, M., & Frazzi, E. (2011). Comparing narrow and broad-band vegetation indices to estimate leaf chlorophyll content in planophile crop canopies. *Precision Agriculture*, *12*(3), 334–344. https://doi.org/10.1007/s11119-010-9204-3

Vohland, M., Mader, S., & Dorigo, W. (2010). Applying different inversion techniques to retrieve stand variables of summer barley with PROSPECT+SAIL. *International Journal of Applied Earth Observation and Geoinformation*, *12*(2), 71–80. https://doi.org/10.1016/j.jag.2009.10.005

Volpato, L., Pinto, F., González-Pérez, L., Thompson, I. G., Borém, A., Reynolds, M., Gérard, B., Molero, G., & Rodrigues, F. A. (2021). High Throughput Field Phenotyping for Plant Height Using UAV-Based RGB Imagery in Wheat Breeding Lines: Feasibility and Validation. *Frontiers in Plant Science*, *12*, 185. https://doi.org/10.3389/fpls.2021.591587

Vuolo, F., Neugebauer, N., Bolognesi, S. F., Atzberger, C., & D'Urso, G. (2013). Estimation of Leaf Area Index Using DEIMOS-1 Data: Application and Transferability of a Semi-Empirical Relationship between two Agricultural Areas. *Remote Sensing*, *5*(3), 1274–1291. https://doi.org/10.3390/rs5031274

Wan, L., Zhang, J., Dong, X., Du, X., Zhu, J., Sun, D., Liu, Y., He, Y., & Cen, H. (2021). Unmanned

aerial vehicle-based field phenotyping of crop biomass using growth traits retrieved from PROSAIL model. *Computers and Electronics in Agriculture*, *187*, 106304. https://doi.org/10.1016/j.compag.2021.106304

Wan, L., Zhu, J., Du, X., Zhang, J., Han, X., Zhou, W., Li, X., Liu, J., Liang, F., He, Y., & Cen, H. (2021). A model for phenotyping crop fractional vegetation cover using imagery from unmanned aerial vehicles. *Journal of Experimental Botany*, *72*(13), 4691–4707. https://doi.org/10.1093/jxb/erab194

Wang, F., Huang, J., Tang, Y., & Wang, X. (2007). New Vegetation Index and Its Application in Estimating Leaf Area Index of Rice. *Rice Science*, *14*(3), 195–203. https://doi.org/10.1016/S1672-6308(07)60027-4

Wang, J., Badenhorst, P., Phelan, A., Pembleton, L., Shi, F., Cogan, N., Spangenberg, G., & Smith, K. (2019). Using Sensors and Unmanned Aircraft Systems for High-Throughput Phenotyping of Biomass in Perennial Ryegrass Breeding Trials. *Frontiers in Plant Science*, *10*. https://doi.org/10.3389/fpls.2019.01381

Wang, Li'ai, Zhou, X., Zhu, X., Dong, Z., & Guo, W. (2016). Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. *The Crop Journal*, *4*(3), 212–219. https://doi.org/10.1016/j.cj.2016.01.008

Wang, Li, Chen, S., Peng, Z., Huang, J., Wang, C., Jiang, H., Zheng, Q., & Li, D. (2021). Phenology Effects on Physically Based Estimation of Paddy Rice Canopy Traits from UAV Hyperspectral Imagery. *Remote Sensing*, *13*(9), 1792. https://doi.org/10.3390/rs13091792

Wang, Y., Wang, D., Shi, P., & Omasa, K. (2014). Estimating rice chlorophyll content and leaf nitrogen concentration with a digital still color camera under natural light. *Plant Methods*, *10*(1), 36. https://doi.org/10.1186/1746-4811-10-36

Warren, C. R. (2008). Rapid Measurement of Chlorophylls with a Microplate Reader. *Journal of Plant Nutrition*, *31*(7), 1321–1332. https://doi.org/10.1080/01904160802135092

White, J. W., Andrade-Sanchez, P., Gore, M. A., Bronson, K. F., Coffelt, T. A., Conley, M. M., Feldmann, K. A., French, A. N., Heun, J. T., Hunsaker, D. J., Jenks, M. A., Kimball, B. A., Roth, R. L., Strand, R. J., Thorp, K. R., Wall, G. W., & Wang, G. (2012). Field-based phenomics for plant genetics research. *Field Crops Research*, *133*, 101–112. https://doi.org/10.1016/j.fcr.2012.04.003

Widlowski, J.-L., Verstraete, M. M., Pinty, B., & Gobron, N. (2000). Advanced vegetation indices optimized for up-coming sensors: Design, performance, and applications. *IEEE Transactions on Geoscience and Remote Sensing*, *38*(6), 2489–2505. https://doi.org/10.1109/36.885197

Wolanin, A., Mateo-García, G., Camps-Valls, G., Gómez-Chova, L., Meroni, M., Duveiller, G., Liangzhi, Y., & Guanter, L. (2020). Estimating and understanding crop yields with explainable deep learning in the Indian Wheat Belt. *Environmental Research Letters*, *15*(2), 024019. https://doi.org/10.1088/1748-9326/ab68ac

Wood, S. N. (2017). *Generalized Additive Models*. Chapman and Hall/CRC. https://doi.org/10.1201/9781315370279

Woźniak, G., Dyderski, M. K., Kompała-Bąba, A., Jagodziński, A. M., Pasierbiński, A., Błońska, A., Bierza, W., Magurno, F., Sierka, E., Kompała-Bąba, A., Jagodziński, A. M., Pasierbiński, A., Błońska, A., Bierza, W., Magurno, F., & Sierka, E. (2021). Use of remote sensing to track postindustrial vegetation development. *Land Degradation & Development*, *32*(3), 1426–1439. https://doi.org/10.1002/ldr.3789

Wu, C., Niu, Z., Tang, Q., & Huang, W. (2008). Estimating chlorophyll content from hyperspectral vegetation indices: Modeling and validation. *Agricultural and Forest Meteorology*, *148*(8–9),

1230–1241. https://doi.org/10.1016/j.agrformet.2008.03.005

Xie, C., & Yang, C. (2020). A review on plant high-throughput phenotyping traits using UAV-based sensors. *Computers and Electronics in Agriculture*, *178*, 105731. https://doi.org/10.1016/j.compag.2020.105731

Xie, Q., Mayes, S., & Sparkes, D. L. (2016). Early anthesis and delayed but fast leaf senescence contribute to individual grain dry matter and water accumulation in wheat. *Field Crops Research*, *187*, 24–34. https://doi.org/10.1016/j.fcr.2015.12.009

Xu, X., Fan, L., Li, Z., Meng, Y., Feng, H., Yang, H., & Xu, B. (2021). Estimating Leaf Nitrogen Content in Corn Based on Information Fusion of Multiple-Sensor Imagery from UAV. *Remote Sensing*, *13*(3), 340. https://doi.org/10.3390/rs13030340

Yang, G., Liu, J., Zhao, C., Li, Z. Z., Huang, Y., Yu, H., Xu, B., Yang, X., Zhu, D., Zhang, X., Zhang, R., Feng, H., Zhao, X., Li, Z. Z., Li, H., & Yang, H. (2017). Unmanned Aerial Vehicle Remote Sensing for Field-Based Crop Phenotyping: Current Status and Perspectives. *Frontiers in Plant Science*, *8*, 1111. https://doi.org/10.3389/fpls.2017.01111

Yang, H. (2011). Remote Sensing Technique for Predicting Harvest Time of Tomatoes. *Procedia Environmental Sciences*, *10*, 666–671. https://doi.org/10.1016/j.proenv.2011.09.107

Yang, J., & Udvardi, M. (2018). Senescence and nitrogen use efficiency in perennial grasses for forage and biofuel production. *Journal of Experimental Botany*, *69*(4), 855–865. https://doi.org/10.1093/jxb/erx241

Yang, Y., Zha, W., Tang, K., Deng, G., Du, G., & Liu, F. (2021). Effect of Nitrogen Supply on Growth and Nitrogen Utilization in Hemp (Cannabis sativa L.). *Agronomy*, *11*(11), 2310. https://doi.org/10.3390/agronomy11112310

Yu, N., Li, L., Schmitz, N., Tian, L. F., Greenberg, J. A., & Diers, B. W. (2016). Development of methods to improve soybean yield estimation and predict plant maturity with an unmanned aerial vehicle based platform. *Remote Sensing of Environment*, *187*, 91–101. https://doi.org/10.1016/j.rse.2016.10.005

Yue, J., Feng, H., Yang, G., & Li, Z. (2018). A Comparison of Regression Techniques for Estimation of Above-Ground Winter Wheat Biomass Using Near-Surface Spectroscopy. *Remote Sensing*, *10*(2), 66. https://doi.org/10.3390/rs10010066

Zaman-Allah, M., Vergara, O., Araus, J. L., Tarekegne, A., Magorokosho, C., Zarco-Tejada, P. J., Hornero, A., Albà, A. H., Das, B., Craufurd, P., Olsen, M., Prasanna, B. M., & Cairns, J. (2015). Unmanned aerial platform-based multi-spectral imaging for field phenotyping of maize. *Plant Methods*, *11*(1), 35. https://doi.org/10.1186/s13007-015-0078-2

Zegada-Lizarazu, W., Elbersen, H. W., Cosentino, S. L., Zatta, A., Alexopoulou, E., & Monti, A. (2010). Agronomic aspects of future energy crops in Europe. *Biofuels, Bioproducts and Biorefining*, *4*(6), 674–691. https://doi.org/10.1002/bbb.242

Zhang, B., Hastings, A., Clifton-Brown, J. C., Jiang, D., & Faaij, A. P. C. (2020). Modeled spatial assessment of biomass productivity and technical potential of Miscanthus × giganteus , Panicum virgatum L., and Jatropha on marginal land in China. *GCB Bioenergy*, *12*(5), 328–345. https://doi.org/10.1111/gcbb.12673

Zhang, C., & Kovacs, J. M. (2012). The application of small unmanned aerial systems for precision agriculture: a review. *Precision Agriculture*, *13*(6), 693–712. https://doi.org/10.1007/s11119-012-9274-5

Zhang, F., & Zhou, G. (2019). Estimation of vegetation water content using hyperspectral vegetation

indices: a comparison of crop water indicators in response to water stress treatments for summer maize. *BMC Ecology*, *19*(1), 18. https://doi.org/10.1186/s12898-019-0233-0

Zhang, M., Abrahao, G., Cohn, A., Campolo, J., & Thompson, S. (2021). A MODIS-based scalable remote sensing method to estimate sowing and harvest dates of soybean crops in Mato Grosso, Brazil. *Heliyon*, *7*(7), e07436. https://doi.org/10.1016/j.heliyon.2021.e07436

Zhang, Y., Yang, J., Liu, X., Du, L., Shi, S., Sun, J., & Chen, B. (2020). Estimation of Multi-Species Leaf Area Index Based on Chinese GF-1 Satellite Data Using Look-Up Table and Gaussian Process Regression Methods. *Sensors*, *20*(9), 2460. https://doi.org/10.3390/s20092460

Zhou, J. J., Yungbluth, D., Vong, C. N., Scaboo, A., & Zhou, J. J. (2019). Estimation of the Maturity Date of Soybean Breeding Lines Using UAV-Based Multispectral Imagery. *Remote Sensing*, *11*(18), 2075. https://doi.org/10.3390/rs11182075

Zhu, W., Sun, Z., Huang, Y., Lai, J., Li, J., Zhang, J., Yang, B., Li, B., Li, S., Zhu, K., Li, Y., & Liao, X. (2019). Improving Field-Scale Wheat LAI Retrieval Based on UAV Remote-Sensing Observations and Optimized VI-LUTs. *Remote Sensing*, *11*(20), 2456. https://doi.org/10.3390/rs11202456

Zhu, X.-G., Chang, T.-G., Song, Q.-F., Finnan, J., Barth, S., Mårtensson, L.-M., & Jones, M. B. (2016). A Systems Approach Guiding Future Biomass Crop Development on Marginal Land. In *Perennial Biomass Crops for a Resource-Constrained World* (pp. 209–224). Springer International Publishing. https://doi.org/10.1007/978-3-319-44530-4_18

Acknowledgments

I would like to thank all the people who inspired, encouraged, and helped me during this memorable PhD journey. Without them, I would have never finished this thesis.

I would like to thank my supervisor prof. Stefano Amaducci for providing me autonomy, useful advice and many helpful suggestions that were invaluable for the development of my research ideas and skills, necessary to complete my PhD.

I would like to thank my colleagues, even those working on other scientific fields, for helping me to gather and process data, and for their friendly support.

I would like to thank "Fondazione Eugenio e Germana Parizzi", in particular Dr. Cinzia Parizzi, for financially supporting my research project.

This thesis 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 from the Bio-based Industries Consortium.

Short biography

Giorgio Impollonia was born on February 19, 1990, in Palermo (Italy). Giorgio studied "Agricultural Science and Technology" at "Università degli Studi di Palermo" in Palermo (Italy), where he obtained his master's degree in 2016, discussing the thesis "Development of a multispectral tool for the determination of turfgrass vegetation indices through imaging techniques". After his master's degree, Giorgio attended a postgraduate on "GIScience and Remote Pilot Systems, for the integrated management of the land and natural resources: Geoinformation and new technologies for sustainable agriculture" at "Università degli Studi di Padova" in Padua (Italy), where he obtained his postgraduate in 2018, discussing the thesis "Development of radiometric correction techniques for UAV remote sensing images for monitoring the putting green turfgrass". During his postgraduate in Padua, Giorgio developed an interest in geo-informatics and remote sensing for precision agriculture, and he collaborated with Archetipo start-up, which provides remote sensing services. In November 2018, Giorgio won a scholarship, allowing him to carry out his PhD at the Centre for Geospatial Research and Analysis and Remote Sensing (CRAST) at "Università Cattolica del Sacro Cuore" in Piacenza (Italy). During his PhD, he worked in remote sensing for precision agriculture, high-throughput crop phenotyping, and participated in several scientific projects (see the section PhD scientific activities). His research field is at the interface between GIS science and remote sensing for precision agriculture applications. His main interests are related to high resolution multispectral/thermal UAV remote sensing to estimate biochemical and biophysical crop parameters, such as leaf area index (LAI), chlorophyll and nitrogen contents, biomass, and yield, through machine learning analysis and inversion of the canopy radiative transfer models.

List of publications

Published papers:

- Motisi, A., **Impollonia**, **G**., Minacapilli, M., Orlando, S., & Sarno, M. (2021). TURF-BOX: an active lighting multispectral imaging system with led VIS-NIR sources for monitoring of vegetated surfaces. In Acta Horticulturae (Issue 1314, pp. 383–390). International Society for Horticultural Science (ISHS). https://doi.org/10.17660/actahortic.2021.1314.48

- **Impollonia**, **G**., Croci, M., Martani, E., Ferrarini, A., Kam, J., Trindade, L. M., Clifton-Brown, J., Amaducci, S. (2022). Moisture content estimation and senescence phenotyping of novel *Miscanthus* hybrids combining UAV based remote sensing and machine learning. GCB Bioenergy. https://doi.org/10.1111/gcbb.12930

Papers in review and in preparation:

- Antonucci[*], G., **Impollonia**, **G**.[*], Croci, M., Marcone, A., Amaducci, S. (2022). High-Throughput Plant Phenotyping for Evaluating Biostimulants: Biophysical Variables Estimation Through PROSAIL Inversion. Submitted to Smart Agricultural Technology. [*] Shared co-first authorship

- **Impollonia**, **G**., Croci, M., Ferrarini, A., Brook, J., Martani, E., Blandinières, H., Marcone, A., Awty-Carroll, D., Ashman, C., Kam, J., Trindade, L. M., Boschetti, M., Clifton-Brown, J., Amaducci, S. (202x). UAV remote sensing for high-throughput phenotyping and for yield prediction of *Miscanthus* by machine learning techniques.

- **Impollonia**, **G**., Croci, M., Blandinières, H., Marcone, A., Amaducci, S. (202x). Comparison of different inversion methods of the PROSAIL model for the hemp trait estimations by UAV-based phenotyping.

- Croci, M., **Impollonia**, **G**., Colauzzi, M., Amaducci, S. (202x). Impact of training set size and lead time on early processing tomato crop mapping accuracy.

- Croci, M., **Impollonia**, **G**., Letterio, T., Marcone, A., Colauzzi, M., Ventura, F., Vignudelli, M., Anconelli, S., Amaducci, S. (202x). Comparative analysis of different retrieval methods for mapping Leaf Area Index using Sentinel-2.

- Croci, M., **Impollonia**, **G**., Amaducci, S. (202x). Dynamic Maize Yield Predictions Using Machine Learning on Multi-Source Data.

- Blandinières, H., **Impollonia**, **G**., Croci, M., Marcone, A., Amaducci, S. (202x). Productivity, eco-physiology, and stem processability of a yellow hemp (*Cannabis sativa* L.) cultivar under varying levels of nitrogen fertilisation.

PhD scientific activities

PhD exams of the doctoral school

- Research ethics and epistemology
- European food law and policy
- English course
- Introduction to applied statistics for agri-food data
- Sustainable animal production
- Information literacy
- Food technology and sustainability
- Accounting tools for management
- Sustainability management
- Ecological intensification of agriculture

Courses

- "Spatial data analysis, applications to agronomic research" organized by Società Italiana di Agronomia.
- "Statistical methodology for agricultural science. General and generalized linear models" organized by Società Italiana di Agronomia.

Projects

Here is a list of the projects that I've participated during the PhD period:

- **Agro.Big.Data.Science:** The Agro.Big.Data.Science project intends to apply the data driven logic to 3 production chains (kiwi, pear and spinach), complete with the necessary sensors for real-time data collection.
- **FarmCO_2Sink:** The FarmCO_2Sink project aims to identify, in agroforestry production systems of the Province of Piacenza, the most efficient conservation systems in terms of biological sequestration of C at farm level and the reduction of climate-altering greenhouse gas emissions (GHG).
- **GRACE:** The EU-funded GRACE project demonstrates large-scale miscanthus and hemp production on land with low productivity, contaminated soil or which has been abandoned.
- **POSITIVE:** The POSITIVE project aims to make agronomic interest indices available on a regional scale from satellite images of the Copernicus project and to prepare an IT infrastructure that makes precision irrigation and fertigation usable throughout the region.

- **SOIPomI:** The SOIPomI project aims to remotely monitor the fields, processing the satellite images of the European Space Agency, to give operators in the supply chain an innovative tool to improve the management of processing tomatoes in Northern Italy.

Lecturer

- "UAV remote sensing for precision nitrogen fertilisation" carried out during the subject "Agricoltura di Precisione" of master's degree "Scienze e Tecnologie Agrarie" of the "Università Cattolica del Sacro Cuore".
- "Case studies of precision agriculture" carried out at the Dinamica training centre.
- "GIS for precision agriculture" carried out at the Dinamica training centre.
- "Introduction to GIS and information systems" carried out at the Dinamica training centre.
- "Precision agriculture for nitrogen fertilisation" carried out at the Formart training centre.
- "Crop irrigation and precision irrigation" carried out at the Vittorio Tadini training centre.
- "GIS for precision agriculture" carried out at the Vittorio Tadini training centre.
- "Precision agriculture" carried out at the Vittorio Tadini training centre.
- "Monitoring of spinach by remote sensing" webinars carried out within Agro.Big.Data.Science project.