

Forschungszentrum Jülich  
IBG 2: Institute for Bio- and Geosciences – Plant  
Sciences

# **UAV remote sensing for the spatial differentiated assessment of plant traits based on multispectral images**

**Dissertation**

zur Erlangung des Grades

Doktor der Agrarwissenschaften (Dr. agr.)

der Landwirtschaftlichen Fakultät

der Rheinischen Friedrich-Wilhelms-Universität Bonn

**Norman Wilke**

aus

Mechernich

Bonn 2023

Angefertigt mit Genehmigung der Landwirtschaftlichen Fakultät der Universität Bonn

Referent: Prof. Dr. Uwe Rascher

Koreferent: Prof. Dr. Ralf Pude

Tag der mündlichen Prüfung: 23.08.2022

## Summary

Rising global temperatures, carbon dioxide (CO<sub>2</sub>) concentration, and the occurrence of extreme weather events are constantly changing the environmental conditions for crop productivity. Identifying climate-adapted and high-yielding cultivars has become a challenging task for the plant breeding community. Time-efficient and objective crop phenotyping is advantageous in reliably identifying desired varieties. Moreover, the identification of variabilities within the agricultural system is an important means of enabling site-specific management and increasing agricultural sustainability. In recent years, unmanned aerial vehicles (UAVs) have opened new frontiers for high-throughput phenotyping. Small-sized and lightweight sensors can be equipped on UAVs, facilitating the mapping of field heterogeneity, with a high spatial and temporal resolution. The enormous potential to assess different crop parameters non-destructively using the three-dimensional (3D) or spectral information has been investigated in this work. In addition, the potential limitations and the effort of applying the developed methods to unknown datasets are examined in this thesis, to meet the needs of precision agriculture and plant breeding.

The first study focuses on the lodging assessment of barley, based on the 3D information. Canopy height models were utilized to quantify the presence or absence of lodged plants (lodging percentage), as well as the intensity of the permanent displacement of plants from their upright position (lodging severity). Height thresholds that make use of a mathematical approach enabled to distinguish between the naturally occurring height variations and the lodged plants. The results showed a high correlation of lodging percentage to reference data (RMSE = 7.66 %) when applied to breeding trials. A similar accuracy could be observed in the case of a farmer's field. The use of the 3D information is nearly independent of the abiotic or biotic factors and is characterized by high repeatability and applicability.

The second study of this thesis makes use of multispectral imagery for the plant density assessment of barley and wheat. The methodology was based on an empirical regression model and the relationship between plant density, counted in the field, and calculated fractional cover (number of plant pixels per square meter). The results illustrated that, at an early seedling stage, fractional cover is closely related to the number of plants (RMSE < 34 plants m<sup>-2</sup>). In order to expose the robustness of this statistical relationship, a developed model was applied to an unknown dataset and showed a high degree of accuracy (RMSE = 24 plants m<sup>-2</sup>). In addition, 11 reference measurements have proved to be sufficient, to enable the adaptation of the empirical regression model to the scene.

The third study explores the application of high-resolution RGB imagery for the spike density assessment of wheat. Similar to the previous study, the statistical relationship between the manually counted spike density and the UAV-derived spike cover (number of spike pixels per square meter) was examined. The spike density could be accurately modelled (RMSE < 18 spikes m<sup>-2</sup>) for three different genotypes, with phenotypic variations of canopy and spike characteristics. Investigations in this study have led to the assumption that 11 reference measurements can enable the adaptation of the empirical regression model.

In summary, the work demonstrated the great diversity of mapping crop heterogeneity with UAVs on a field scale. This potentially enables the faster selection of superior lines, the spatially differentiated prediction of crop yield or site-specific management for different nutrition and soil conditions. The independency or known effort of adapting the methodologies to unknown datasets increases the application potential in practice. However, the greatest challenge in remote sensing is to demonstrate universal applicability or the satisfactory adaptation of the methods.

## Zusammenfassung

Der Anstieg der globalen Temperaturen, der Kohlendioxid (CO<sub>2</sub>) Konzentration sowie das Auftreten extremer Wetterereignisse verändern die Umweltbedingungen für die Produktivität von Nutzpflanzen kontinuierlich. Die Identifizierung von klimaangepassten und ertragreichen Sorten ist für die Pflanzenzüchtung zu einer anspruchsvollen Aufgabe geworden. Eine zeiteffiziente und objektive Phänotypisierung von Nutzpflanzen ist dabei vorteilhaft, um die richtigen Sorten zuverlässig identifizieren zu können. Zusätzlich ist die Identifizierung der Variabilität innerhalb eines landwirtschaftlichen Systems wichtig, um standortspezifisches Management zu ermöglichen und die Nachhaltigkeit in der Landwirtschaft zu erhöhen. In den letzten Jahren haben Unmanned Aerial Vehicles (UAVs) neue Möglichkeiten für die Hochdurchsatz-Phänotypisierung eröffnet. UAVs können mit kleinen und leichten Sensoren ausgestattet werden, die es ermöglichen, Heterogenitäten mit einer hohen räumlichen und zeitlichen Auflösung zu erfassen. Dieses Potenzial wird in der vorliegenden Arbeit untersucht, um verschiedene Pflanzenparameter nicht destruktiv und anhand dreidimensionaler (3D) oder spektraler Informationen zu quantifizieren. Darüber hinaus werden in dieser Arbeit mögliche Grenzen sowie der Aufwand für die Anwendung der entwickelten Methoden auf unbekannte Datensätze untersucht, um den Anforderungen der Präzisionslandwirtschaft und der Pflanzenzüchtung gerecht zu werden.

Dabei befasst sich die erste Studie mit der Ermittlung der Lagerbildung bei Gerste auf Grundlage der 3D-Information. Mit Hilfe eines Oberflächenmodelles von Pflanzenbeständen wird die Betroffenheit und nicht Betroffenheit von Pflanzen im Bezug zur Lagerbildung (Lageranteil) sowie deren Intensität (Lagerintensität) quantifiziert. Hörschwellwerte, die auf einem mathematischen Ansatz beruhen, ermöglichen die Unterscheidung zwischen natürlich vorkommenden Hörschwankungen und der Lagerbildung. Die Ergebnisse zeigen eine hohe Korrelation des ermittelten Lageranteils zu Referenzdaten (RMSE = 7,66 %) bei Zuchtparzellen. Eine ähnliche Genauigkeit konnte auch für eine klassische landwirtschaftliche genutzte Fläche erzielt werden. Die Nutzung der 3D-Informationen ist nahezu unabhängig von abiotischen und biotischen Faktoren und zeichnet sich durch eine hohe Wiederholbarkeit und Anwendbarkeit aus.

Im zweiten Teil dieser Arbeit werden multispektrale Bilder zur Ermittlung der Pflanzendichte von Gerste und Weizen genutzt. Die Methodik basiert auf einem empirischen Regressionsmodell und der Beziehung zwischen der auf dem Feld gezählten Pflanzendichte und dem berechneten Bedeckungsgrad (Anzahl Pflanzenpixel pro Quadratmeter). Die Ergebnisse zeigen, dass der Bedeckungsgrad in einem frühen Entwicklungsstadium stark mit der Pflanzendichte korreliert (RMSE < 34 Pflanzen m<sup>-2</sup>). Um die Robustheit dieser statistischen Beziehung aufzuzeigen, wurde ein entwickeltes Modell auf einen unbekannt Datensatz und unbekannt Genotypen mit einer hohen Genauigkeit angewandt (RMSE = 24 Pflanzen m<sup>-2</sup>). Außerdem erwiesen sich bereits elf Referenzmessungen als ausreichend, um das empirischen Regressionsmodell an eine spezifische Situation anzupassen.

Die dritte Studie befasst sich mit der Nutzung von hochauflösenden RGB-Bildern zur Quantifizierung von Ähren in Weizenbeständen. Ähnlich wie in der vorangegangenen Studie wurde die statistische Beziehung zwischen der manuell gezählten Ährendichte und des Ährenbedeckungsgrads (Anzahl Ährenpixel pro Quadratmeter) untersucht. Die Ährendichte konnte für drei verschiedene Genotypen mit phänotypischen Variationen von Kronen- und Ährenmerkmalen genau modelliert werden (RMSE < 18 Ähren m<sup>-2</sup>). Eine Untersuchung der Studie führte ebenfalls zu der Annahme, dass elf Referenzmessungen eine Anpassung des empirischen Regressionsmodells ermöglichen können.

Zusammenfassend lässt sich sagen, dass die Arbeit eine Vielzahl an Möglichkeiten aufzeigt, um Heterogenitäten im Feld mit UAVs erfassen zu können. Dies ermöglicht potenziell eine schnellere Selektion von idealen Züchtungslinien, eine räumlich differenzierte Vorhersage von Ernteerträgen oder eine standortspezifische Bewirtschaftung bei unterschiedlichen Nährstoff- und Bodenbedingungen. Die Unabhängigkeit beziehungsweise der bekannte Aufwand zur Anpassung der Methodik an unbekannt Datensätze erhöht das Anwendungspotenzial in der Praxis. Allerdings ist die größte Herausforderung im Bereich der Fernerkundung die universelle Anwendbarkeit beziehungsweise erfolgsversprechende Anpassung der Methoden nachzuweisen.



## Abbreviations

AGL	above ground level
ALS	average lodging severity
BBCH	Biologische Bundesanstalt, Bundessortenamt und Chemische Industrie
CH	canopy height
CHM	canopy height model
CNN	convolutional neural network
CO <sub>2</sub>	carbon dioxide
DAS	days after sowing
DEM	digital elevation model
DLS	downwelling light sensor
DN	digital number
DTM	digital terrain models
ELM	empirical line method
ExGR	excess green minus excess red
FA	flight altitude
GCP	ground control point
GNSS	global navigation satellite system
GPS	global positioning system
GSD	ground sampling distance
LAI	leaf area index
LiDAR	light detection and ranging
LPT	lodging percentage threshold
MAE	mean absolute error
MAXCH	maximum canopy height
NDVI	normalized difference vegetation index
NIR	near infrared
PA	precision agriculture
RGB	red green blue
rRMSE	relative root mean square error
RMSE	root mean square error
RTK	real time kinematic
R <sup>2</sup>	averaged coefficient of determination (R <sup>2</sup> )
S	Direct solar illumination during data acquisition
(S)	Minor clouds but predominantly direct solar illumination
SD	standard deviation

**Abbreviations continued...**

## IV

SfM	structure from motion
SU	sampling unit
SVM	support vector machines
O	Overcast sky
(O)	Short periods of direct solar illumination but mainly overcast sky
UAV	unmanned aerial vehicle
WALS	weighted average lodging severity
2D	two-dimensional
3D	three-dimensional

## Table of contents

<b>1. Introduction</b>	<b>- 1 -</b>
<b>2. Basics and Methods</b>	<b>- 5 -</b>
2.1 Empirical-statistical regression methods	- 5 -
2.2 Study site	- 6 -
2.3 Remote Sensing with UAVs	- 7 -
2.3.1 UAV system	- 9 -
2.3.2 UAV sensors	- 10 -
2.3.3 Improvements in the UAV setup	- 11 -
2.3.4 Data acquisition, processing, and georeferencing	- 12 -
2.3.5 Calculating reflectance	- 16 -
<b>3. Study I: Quantifying lodging percentage and lodging severity using a UAV-based canopy height model combined with an objective threshold approach</b>	<b>- 17 -</b>
3.1 Introduction	- 18 -
3.2 Materials and methods	- 20 -
3.2.1 Study area	- 20 -
3.2.2 Data acquisition, processing, and canopy height model generation	- 21 -
3.2.3 UAV canopy height assessment and validation	- 22 -
3.2.4 Lodging assessment and validation	- 23 -
3.2.4.1 Experimental site 1: Breeding trials	- 23 -
3.2.4.2 Experimental site 2: Farmer field	- 25 -
3.3 Results and analysis	- 25 -
3.3.1 Comparison of plant traits derived from UAV- and DTM-based ground models	- 25 -
3.3.2 UAV canopy height assessment and validation	- 26 -
3.3.3 Lodging assessment and validation	- 27 -
3.2.4.1 Experimental site 1: Breeding trials	- 23 -
3.2.4.2 Experimental site 2: Farmer field	- 25 -
3.4 Discussion	- 33 -
3.5 Conclusions	- 36 -

<b>4. Study II: Assessment of plant density for barley and wheat using UAV multispectral imagery for high-throughput field phenotyping</b>	<b>- 37 -</b>
4.1 Introduction	- 38 -
4.2 Material and methods	- 40 -
4.2.1 Study site and experimental design	- 40 -
4.2.2 Field validation	- 42 -
4.2.3 UAV data acquisition	- 42 -
4.2.4 Data analysing	- 44 -
4.2.5 Statistical analysis	- 45 -
4.3 Results	- 47 -
4.3.1 Experimental field 1: Barley	- 47 -
4.3.2 Experimental field 2: Wheat	- 48 -
4.3.3 Ideal parameters for data acquisition	- 50 -
4.3.4 Experimental field 3: Investigate repeatability	- 51 -
4.4 Discussion	- 54 -
4.5 Conclusions	- 57 -
<b>5. Study III: Spike density quantification of different wheat varieties from UAV RGB imagery</b>	<b>- 58 -</b>
5.1 Introduction	- 58 -
5.2 Material and methods	- 60 -
5.2.1 Study site and experimental design	- 61 -
5.2.2 UAV data acquisition	- 61 -
5.2.3 Acquisition of the training data set	- 64 -
5.2.4 Spike Segmentation in the extended CIELAB color space	- 64 -
5.2.5 Statistical analysis	- 66 -
5.3 Results	- 67 -
5.3.1 Tybalt	- 67 -
5.3.2 Kadrij	- 69 -
5.3.3 Quintus	- 69 -
5.3.4 Power analysis and stepwise reduction of the sample size	- 71 -
5.4 Discussion	- 73 -
5.5 Conclusions	- 76 -
<b>6. Synthesis</b>	<b>- 77 -</b>
6.1 Thresholding	- 77 -
6.2 Georeferencing	- 78 -

## VII

6.3 Reflectance calculation	- 79 -
6.2 Empirical-statistical regression methods	- 80 -
6.3 Hardware and software requirements	- 81 -
<b>7. Conclusion and outlook</b>	<b>- 83 -</b>
<b>8. Publications</b>	<b>- 85 -</b>
<b>9. References</b>	<b>- 85 -</b>
<b>10. Supplementary materials</b>	<b>- 102 -</b>
<b>11. Acknowledgements</b>	<b>- 102 -</b>

## 1. Introduction <sup>1</sup>

Global agricultural productivity faces enormous challenges. With anticipated growth of the world population to more than nine billion people by the middle of the 21<sup>st</sup> century (United Nations 2015), it will be necessary to nearly double global crop production to meet the food demand (Tilman et al. 2011). In addition, intense land competition (e.g., with urban expansion), increasing meat consumption, and rising demand for bioenergy production put further pressure on global agriculture (Ainsworth et al. 2008; Rathmann et al. 2010; Harvey and Pilgrim 2011; Henchion et al. 2014).

In recent decades, advanced crop breeding and optimization in land use for example, with artificial fertilizer and pesticides have been the main drivers of efforts to meet the steadily increasing food demand (Evenson and Gollin 2003; Pingali 2012). Currently, more than half of the world population is dependent on nitrogen fertilizers for food production (Erismann et al. 2008; Basso et al. 2016). In addition, digestibility has been improved in high-yielding breeds, allowing an increase in average daily gains from livestock (Casler and Vogel 1999). However, current estimates indicate that approximately 690 million people, 9 % of the current world population, are still starving (FAO et al. 2020). Furthermore, the environment in which crops are grown is changing, influencing the productivity of already established crops. The last decade was the warmest since the 19<sup>th</sup> century and led to record-breaking heatwaves in many areas of the world (Coumou et al. 2013; Schiermeier 2019). Increasing temperatures and water stress resulted in a reduction in plant productivity and performance (Challinor et al. 2014; Asseng et al. 2015; Gao et al. 2020). Moreover, carbon dioxide (CO<sub>2</sub>) concentrations are continuously rising. Atmospheric emissions of CO<sub>2</sub> will at least double by 2050 (Houghton et al. 2001; Solomon et al. 2007) with a positive effect on photosynthesis and water-use efficiency. This could be the basis for efforts to offset losses caused by increased drought and temperature stress, particularly for important C3 crops like cereals (Ittersum et al. 2003; Ainsworth et al. 2008; Alexandratos and Bruinsma 2012).

It has become a priority to determine resilient as well as high-yielding crop genotypes under these changing environmental conditions in order to maintain steady gains in yield and ensure global food security. Advanced phenotyping can address these challenges and has significant potential in terms of identifying uncertainties and variabilities within agricultural or breeding systems. This identification is also an important means for precision agriculture (PA), to enable site-specific management and increase crop performance. Tekin (2010) estimated that by applying variable rates of fertilizer, wheat yield could be increased by up to 10 %, and the use of artificial fertilizer could be reduced by up to 37 %. Such an approach would also save valuable resources and reduce negative environmental impacts, for instance, on water and air pollution caused by excessive release of fertilizers or emissions into the atmosphere (Gebbers and Adamchuk 2010; Hunt et al. 2012; Basso et al. 2016).

---

<sup>1</sup> Parts of this section have previously been published by Wilke et al. (2019) and Wilke et al. (2021)

Moreover, advanced phenotyping offers an opportunity to assess the performance of agricultural chemicals, hybrids, and seed types within agricultural or breeding systems under different environmental conditions (Tellaeche et al. 2008; Tekin 2010; Araus and Cairns 2014).

Evaluating, for example, application of chemicals or fertilizers at the right time and place requires qualitative and quantitative information within a short period, to guide the decision-making process. Therefore, it is important to replace the current practice of subjective, time-consuming, and expensive manual surveys of subsample areas to represent field or plot conditions. Remote sensing data from satellites, aircrafts, and UAVs can help to advance and accelerate phenotyping to complement the work of breeders and farmers. These technologies have the ability to provide spectrally, spatially and temporally extensive information about large areas quickly, cost-effectively, and easily (Liaghat and Balasundram 2010; Basso et al. 2016; Weiss et al. 2020).

In particular, with rapid development of sensor technology and data processing, UAVs contribute to increasing digitalization in agriculture and have the potential to change crop phenotyping and agricultural crop production in the coming years (C. Zhang and Kovacs 2012; G. Yang et al. 2017a; Vargas et al. 2020). UAVs are capable of acquiring high-resolution data, with ground sampling distances (GSDs) of centimeters to millimeters, making them ideal for mapping the spatial variability of plants in small-scale experimental plots. Furthermore, due to their flexibility, low cost and simple, safe handling, UAVs have been widely adopted by many other research communities, such as search and rescue (Scherer et al. 2015; Shakhathreh et al. 2019) or in geodesy (Berteška and Ruzgienė 2013; Burdziakowski 2018). Moreover, UAVs are ideal for forestry operations, to monitor fire and its effects (Ollero et al. 2006), to classify species and map diseases (Torresan et al. 2017) or to monitor wildlife (Gonzalez et al. 2016; Ward et al. 2016). The preparation time for UAV campaigns is short with a high flexibility. In addition, data acquisition is not limited by different illumination conditions, and the GSD can be adjusted to the needs of the research, which makes UAVs advantageous for satellite and airborne systems (Zhang and Kovacs 2012; Mulla 2013; Candiago et al. 2015).

These properties of UAVs also facilitate a broad range of agricultural applications, making them suitable for assessing, as example, plant diseases (Nebiker et al. 2016; Abdulridha et al. 2020), plant resistance (Joalland et al. 2018; Chivasa et al. 2020), and plant vigor (Candiago et al. 2015; Wahab et al. 2018). Moreover, various plant traits like the leaf area index (Tian et al. 2017; Comba et al. 2020), canopy height (Holman et al. 2016; Mohan et al. 2017) and biomass (Bendig et al. 2014; Lu et al. 2019) have successfully been quantified non-invasively, to name just a few. The monitoring of chosen variables on a field scale can be used to determine specific heterogeneity time within a breeding trials site or to reactively optimize the micro-management and treat each field section individually as required (Mulla 2013; Srbinovska et al. 2015; Thomasson et al. 2018). Furthermore, information can be used for spatially differentiated predictions of crop yield (Nebiker et al. 2016; Wahab et al. 2018; Hassan et al. 2019).

The two main sources used in the aforementioned studies are 3D information and transformation of certain color bands into a meaningful vegetation index or color space. Through overlapping of acquired images, data were processed using the “structure from motion” (SfM) technique. This enabled calculation of photogrammetric products like 3D point clouds and orthomosaics. The latter allows analysis of a single georeferenced mosaicked image, replacing individual images covering only a small area of interest (Westoby et al. 2012; Thomasson et al. 2018).

Similarly, in this thesis, use was made of transformation of certain color bands into a meaningful vegetation index or color space and 3D point clouds, as sources for method development aimed at quantifying the lodging, plant and spike density of cereals. A UAV was used as the image acquisition platform, taking into account guidelines for qualitative data recording, data processing, and the right timing for trait extraction in relation to the developmental stage. In addition, the influence of varying GSDs for trait extraction was evaluated, which is correlated to the flight altitude of the UAV as well as individual camera parameters. Moreover, in this study, an attempt was made to investigate the accuracy, repeatability and practicability of the method by applying the approach, for example, to unknown datasets. Thus, the accuracy of the developed UAV methodologies was validated with manual reference measurements.

In summary, the main parts of this thesis are structured as follows:

**Section 2** provides an overview about the study site, the basic principles, and applied concepts.

**Section 3** refers to use of 3D point clouds for canopy height extractions to quantify lodging of barley, based on two different parameters for breeding trials and a classical farmer’s field. The first parameter was the lodging percentage, which determined the presence or absence of lodged plants. The second parameter was lodging severity, which takes into account the intensity of permanent displacement of plants from their upright position. The focus was on developing a method that could objectively distinguish between naturally occurring height variations and lodged plants. In addition, the methodology and the factors that influence canopy height extractions were examined.

**Section 4** investigates the relationship between plants counted in the field and fractional cover (number of plant pixels per unit of ground area), determined by two vegetation indices (NDVI and ExGR), derived from UAV-based multispectral imagery. Measurements derived from manual counting in the field are advantageous in that it is possible to intervene in the plant structure or change the view for a better estimation of plant density. The relevant question is how much information can be obtained from nadir UAV images for the respective plant trait.



**Section 5** examines the relationship between the spike density of wheat counted in the image and spike cover (number of spike pixels per square meter), determined by the extended CIELAB color space of spatially highly resolved information in the visible electromagnetic spectrum.

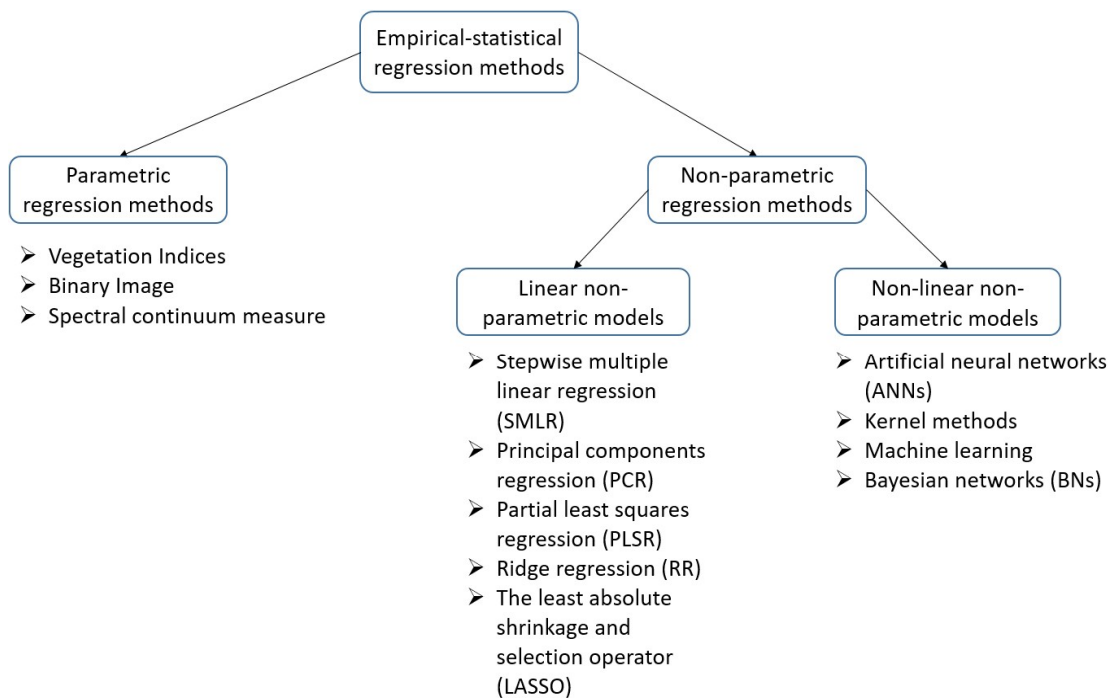
**Section 6** evaluates the advantages and applicability of developed approaches in practice and discusses the limitations of having, for example, different measurement conditions.

**Section 7** summarizes the benefits and future challenges of UAV applications to meet the demands of precision agriculture and breeding communities.

## 2. Basics and methods

### 2.1 Empirical-statistical regression methods

Empirical-statistical regression methods are very popular approaches for predicting biogeophysical vegetation traits, based on spectral information from remote sensing imagery (Dorigo et al. 2007; Verrelst et al. 2015). They can be separated into parametric and non-parametric regression methods (Figure 1). Linear non-parametric models have been extended by machine learning algorithms that go beyond linear transformations. The different approaches for each category are summarized in Figure 1.

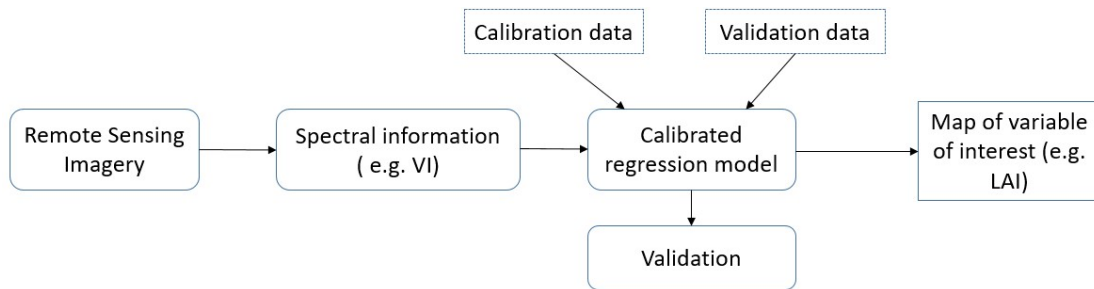


**Figure 1** Overview of different empirical-statistical regression methods (modified according to Verrelst (2015)).

Non-linear transformations with kernel methods like support vector machines (Liakos et al. 2018) and neural networks (Kamilaris and Prenafeta-Boldú 2018) have already shown high potential for trait extractions. Information on, for example, texture, color, morphology and spectrum is used to identify complex associations between predicting variables and the variable of interest. This procedure does not necessitate preselection of data and explicit awareness of the underlying process (Houborg and McCabe 2018; Tewes 2018), and it is especially advantageous for automating trait extractions. However, diversity is critical, and training datasets from thousands to millions are normally required to show significant advantages (Deng et al. 2009; Ubbens et al. 2018; Sadeghi-Tehran et al. 2019). In addition, the training process can be computationally intensive, needs expert knowledge and is potentially adapted to genotypic specificities or illumination conditions with a

restricted level of generalization and applicability. This is reinforced by lack of explicit awareness of the underlying process and by the fact that most methods are considered as block boxes (Verrelst et al. 2015; Morota et al. 2018).

Parametric regression methods just assume a simple relationship between two sets of variables: a dependent variable (variable of interest), such as the leaf area index (LAI), and an independent variable, for example, a spectral transformation like a vegetation index (Figure 2). Prediction of the plant parameter is based on a fitting function between the independent and dependent variable (Verrelst et al. 2015). In this study, parametric regression methods were used for plant and spike density quantification.



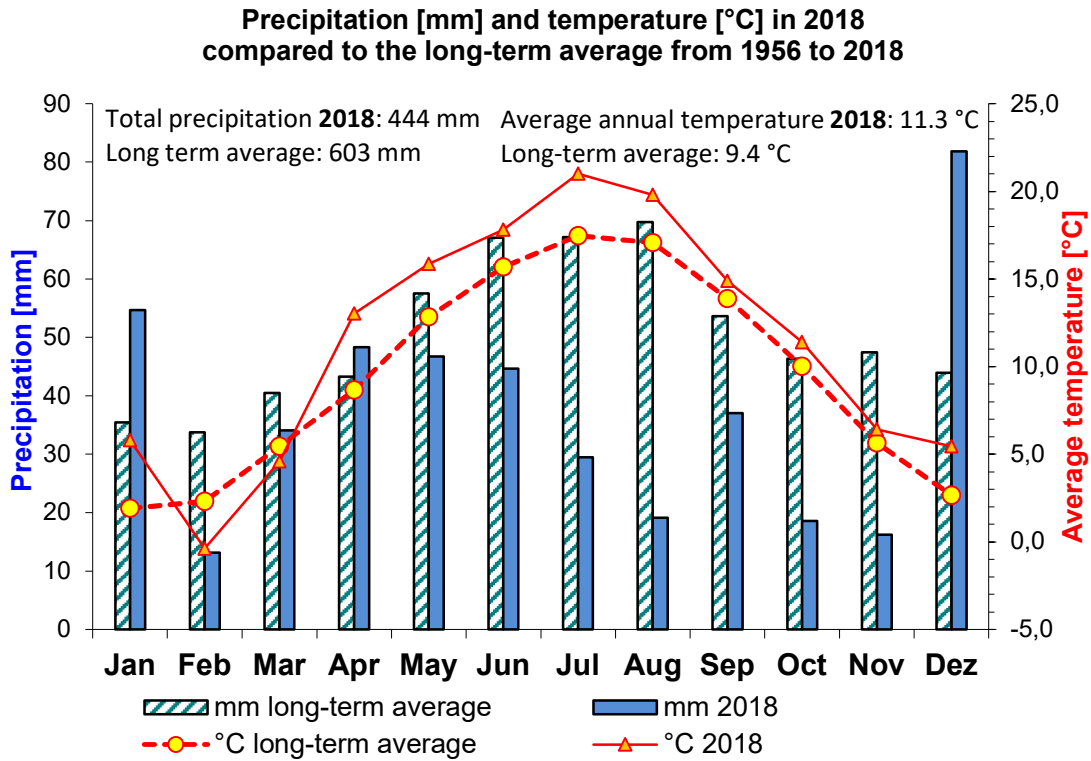
**Figure 2** Generalized flowchart of plant trait retrieval using parametric regression methods (modified according to Verrelst (2015)).

Such statistical relationships can suffer, for example, from a lack of robustness and transferability as they are species-, time-, and site-specific (Homolová et al. 2013). However, in this study, an attempt was made to evaluate the effort required to recalibrate the model. Recalibration does not require expert knowledge and is computationally fast. Respective disadvantages and advantages of non-parametric and parametric regression methods for plant trait identification are discussed in the corresponding subsection of this thesis.

## 2.2 Study site

The experiments were conducted at the agricultural research station Campus Klein-Altendorf, 15 km southwest of the city of Bonn, in part of the Lower Rhine valley. The soil type can be classified as luvisol. It is base- and nutrient-rich and ideal for growing wheat, barley, and sugar beet. Based on the German system of measured productivity (Ackerzahl), the fields have scores between 85 and 90 points, the maximum being 100 points ([www.cka.uni-bonn.de/standort](http://www.cka.uni-bonn.de/standort), accessed January 02, 2022). The study area is located on the lee side of the Eifel and is dominated by westerly winds. The weather station of the campus, in operation since 1956, has reported a long-term average annual temperature of 9.4 °C. The climate is characterized by significant precipitation throughout the year, with an average annual precipitation of 603 mm (Kunz and Völkerling 2015). The temperatures and amount of precipitation in the long-term monthly averages are illustrated in Figure 3, in comparison to the year 2018. This year was selected because it was the essential period for plant (Section 4) and spike (Section 5) density assessment in this thesis.

In 2018, the average annual precipitation was 150 mm lower, and the average annual temperature was around 2 °C higher, compared to the long-term average. For the investigated spring cultivars of barley and wheat, the temperatures in April (up to 25 °C, well above the daily average) and precipitation higher than the long-term average led to ideal conditions for plant germination (Figure 3).



**Figure 3** Long-term average climate diagram of precipitation and temperature for Klein-Altendorf in comparison to the year 2018 (an essential period for data acquisition in this thesis).

During spike development from May to July, the temperature was around 3 °C higher and precipitation was lower, compared to the long-term average (Figure 3). Drought and heat stress can influence crop growth and development, leading, for example, to an earlier onset of plant senescence (Samarah 2005; Prasad et al. 2015).



### 2.3 Remote sensing with UAVs

Acquisition of nadir image data from UAVs is highly interesting and helps to obtain information that is only partially visible from the ground. Nowadays, influenced by advances in technology, numerous sensors are available on the market that can be mounted on UAVs. Small-sized thermal image systems have been used to detect, for instance, wildlife and plant water stress (Ward et al. 2016; Messina and Modica 2020). In addition, lightweight multispectral and hyperspectral sensors can deliver information on plant diseases and plant vigor (Nebiker et al. 2016; Wahab et al. 2018; Abdulridha et al. 2020). Moreover, red, green, blue (RGB) cameras can be used to generate, for instance, 3D point clouds, comparable to

light detection and ranging (LiDAR) sensors for height and biomass assessment (Madec et al. 2017; Xie and Yang 2020).

Furthermore, UAV platforms have been broadly diversified and nowadays can fall into two main categories: rotary-wing and fixed-wing UAVs. Both types have their advantages and disadvantages (cf. Table 1). Therefore, selection of a UAV platform must be tailored to the target and the specific field of application. The important features for trait extraction in this study, essential for spike and plant density estimation, are highlighted in bold letters.

**Table 1** Overview of advantages and disadvantages of rotary-wing and fixed-wing UAVs. Important features for data acquisition in this study are highlighted in bold letters (modified according to Dronedeploy (2017)).

		
	Rotary-Wing UAVs	Fixed-Wing UAVs
<b>Maneuverability</b>	✓	☒
Size/portability	✓	☒
<b>Ease of use</b>	✓	☒
Payload capacity	✓	☒
<b>Actively stabilized system for sensor (Gimbal)</b>	✓	☒
<b>Low flight speed</b>	✓	☒
Flight time	☒	✓
Stability	☒	✓
Safer recovery from motor power loss	☒	✓
Efficiency for area mapping	☒	✓

Based on assessment of these features, a rotary-wing UAV was deemed preferable for data acquisition in this study. Essentially, the architecture of fixed-wing UAVs ensures mainly longer flight times with a similar payload. In particular, the maneuverability and flight speed of fixed-wing UAVs do not facilitate recording of high-resolution data with a GSD in the range of millimeters, which was a prerequisite in this thesis. Furthermore, rotary-wing UAVs have an actively stabilized system like a gimbal to correct orientation of the sensor to the UAV movements, which is important in order to avoid motion blur (Gülch 2012; Sieberth et al. 2014).

### 2.3.1 UAV system <sup>2</sup>

The off-the-shelf Falcon 8 octocopter (Ascending Technologies GmbH, Krailing, Germany) was used for data acquisition in this study, which has already proved to be a reliable platform for different scientific purposes (Bláha et al. 2012; Burkart et al. 2015). This small rotary-wing UAV weighs 1700 g and has a maximum payload of 500 g. With a take-off weight of less than 2 kg, there was no need to have an operating license according to German regulations until 2020. In addition, this lightweight UAV generates little wind, which is advantageous for proximal sensing. A single 6.250-mAh lithium polymer battery powers the system and enables a flight duration of 10 to 15 minutes. It is mainly wind speed, payload and the number of charging cycles of the battery that influence variations in flight duration. The platform was fully designed and built by the manufacturer, with a global positioning system (GPS) receiver and autopilot. This guaranteed simple handling and reliable flight conditions for autonomous image acquisition. The patented V-shaped design of the UAV, illustrated in Figure 4, enables integration of sensors into a gimbal that looks upwards and downwards, without the sensors' field of view being blocked by other components of the UAV.



**Figure 4** Falcon 8 octocopter used for data acquisition in the presented study (image provided by Ralf-Uwe Limbach).

The gimbal is an important component which compensates for pitch and roll movement of the UAV platform during data acquisition and guarantees that the sensor has a stable nadir view during data acquisition. However, space and payload are limited on the Falcon 8, with the need to exchange the setup by different sensors.

---

<sup>2</sup> Parts of this section have previously been published by Wilke et al. (2019) and Wilke et al. (2021)

### 2.3.2 UAV sensors <sup>3</sup>

In this study, the UAV system was equipped with two different sensors: a consumer-grade digital RGB Sony Alpha 6000 (Sony Europe Limited, Weybridge, Surrey, UK) camera and the MicaSense RedEdge-3 multispectral sensor (Micasense, Seattle, USA). The former has a resolution of 24 megapixels (6000 × 4000 pixels) and captures images in the visible wavelength range (400 – 700 nm). The UAV manufacturer integrated the sensor into the Falcon 8 system (Figure 4), which allowed live transmission to the controller and the possibility of triggering the camera manually. The GPS receiver and autopilot of the Falcon 8 can also be used to trigger the camera, based on a waypoint pattern with intended across- and along-track overlap. The RGB images were not geotagged. Nadir images were acquired in JPEG format and sRGB color standard (Süsstrunk et al. 1999) at a shutter speed of 1/1000 seconds.

The multispectral MicaSense sensor has already proven its effectiveness in different agricultural studies (Walsh et al. 2018; Gong et al. 2019). The global shutter simultaneously captures five spectral bands in 12-bit RAW TIFF format with GPS information. Details of the center wavelength of each band and the associated bandwidth from blue to near infrared (NIR) are summarized in Table 2.

**Table 2** Spectral characteristics of the MicaSense RedEdge-3.

<b>Band</b>	<b>Center Wavelength (nm)</b>	<b>Bandwidth (nm)</b>
Blue	475	32
Green	560	27
Red	668	14
Red edge	717	12
NIR	842	57

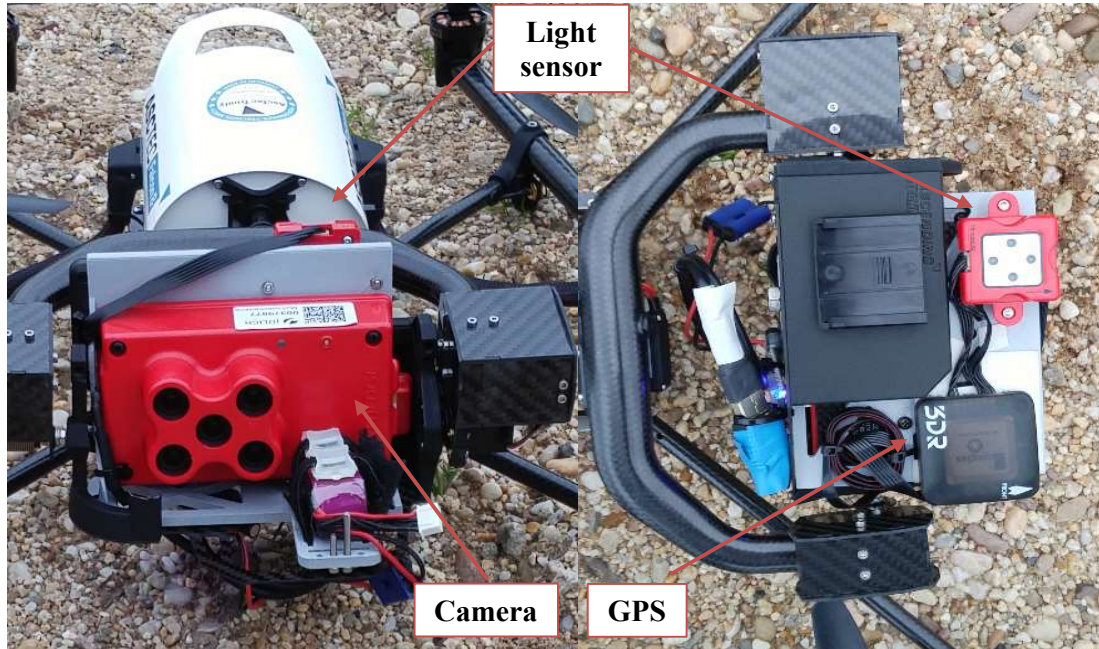
In terms of weight (150 g) and dimensions (12.1 cm x 6.6 cm x 4.6 cm), the sensor was ideal for the available UAV platform. The capture rate of 1 second per image enabled data acquisition from low flight altitudes and at high flight speed. Exposure and gain settings are automatically optimized for each captured image and spectral band to avoid blurring or overexposure, resulting in correctly exposed images (Walsh et al. 2018). The images of the five spectral bands acquired by the multispectral camera were separately saved as TIFF files. An additional important feature of the sensor is the downwelling light sensor (DLS), which records the downwelling incoming light (irradiance) for each spectral band during a flight. Moreover, the camera manufacturer provides users with a radiometrically calibrated reference panel.

---

<sup>3</sup> Parts of this section have previously been published by Wilke et al. (2019) and Wilke et al. (2021)



The setup of the MicaSense camera was integrated into the Falcon 8 system. The patented V-shaped design of the UAV enabled the DLS to be placed on the back of the gimbal, facing towards the sky (Figure 5). Therefore, both the camera and the DSL have an identical correction of orientation to the UAV's movements.



**Figure 5** MicaSense multispectral camera integrated into the Falcon 8 system (left), including a GPS module and light sensor mounted on the back of the gimbal (right).

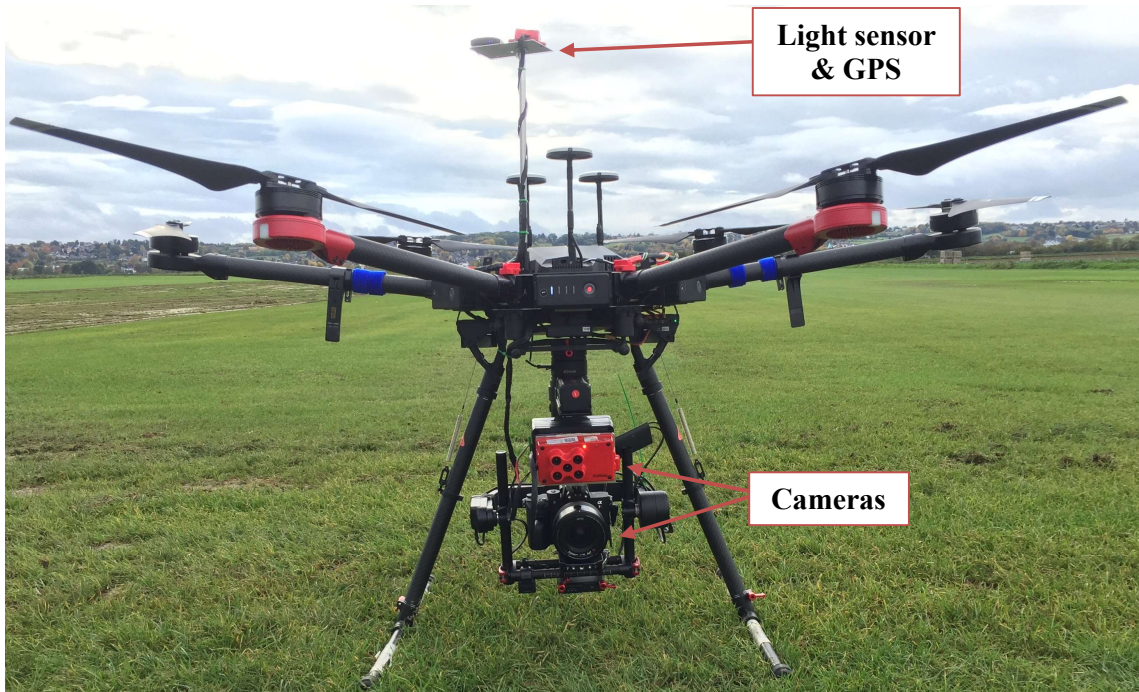
The weight of the DLS, GPS, camera, mounting plate and power supply battery was around 230 g, resulting in a total UAV take-off weight of less than 2 kg.

### 2.3.3 Improvements in the UAV setup

The UAV market has been driven by innovations as well as price decreases over the past few years. In comparison to other systems available on the market, the flight time and payload of the Falcon 8 UAV are clearly limited, leaving room for improvement in terms of more efficient data acquisition. The main criteria in this study were simultaneous data acquisition with a high-resolution RGB and multispectral sensor, a higher payload and an increase in flight time. Bearing in mind this list of requirements, the DJI Matrice 600 Pro Hexacopter with a Ronin MX gimbal, which has already proved to be a reliable platform in other scientific studies (Horstrand et al. 2019; Tait et al. 2019), seemed a good choice. The system also enables simple handling and stable and reliable autonomous flight conditions. With a total weight of 12.5 kg, including the gimbal, the Hexacopter requires a drone license for operation.



The RGB and the multispectral sensors were attached to the gimbal for simultaneous data acquisition (Figure 6). The common drone architecture enabled DSL integration only on the top of the UAV, without correction of sensor orientation to UAV's movements during data acquisition.



**Figure 6** DJI Matrice 600 Pro and Ronin MX gimbal, with integrated high-resolution RGB camera and MicaSense multispectral sensor

As with the Falcon 8 UAV system, the company offers live transmission and manual triggering of the RGB camera. The multispectral camera with the smaller field of view of the two cameras determined the UAV waypoint pattern to ensure sufficient across- and along-track overlap. This setup and six 5.700-mAh lithium polymer batteries enabled a flight time of more than 20 minutes.

### 2.3.4 Data acquisition, processing, and georeferencing <sup>4</sup>

The photogrammetry software Agisoft Professional (Agisoft LLC, Saint Petersburg, Russia) makes use of the SfM technique to process single UAV images. Through a defined number of overlapping images, the algorithms identify corresponding images by means of feature recognition. Details of the algorithms can be found in several publications (Lowe 2004; Kersten 2016; Agisoft 2018). In this study, images were acquired with 60 % across- and 80 % along- track overlap in the lawnmower pattern (Dos Santos et al. 2015). This is supported by the software manufacturer's recommendations to avoid image artifacts and to enable reliable reconstruction of the scene (Agisoft 2018), which has been successfully applied in several studies (Colomina and Molina 2014; Soares et al. 2018). Based on these overlapping images, the software recreates their final orientation in a spatial three-dimensional (3D) point cloud (Özyeşil et al. 2017). Reconstruction of the 3D point cloud is an important prerequisite for generation of a geometrically corrected mosaicked image of the entire scene.

The photogrammetric process can be divided into the following five steps:

1. **Align photos:** The alignment process identifies prominent points within an image to detect similarities and matches across the image sequences. Thus, a sparse point cloud is calculated based on identified corresponding points in different images. This sparse point cloud can already be used to manually add ground control points (GCPs) with known locations to georeference the product in a global coordinate system (WGS 84 / UTM Zone 32N). The exact locations of the GCPs were measured using a Trimble differential real-time kinematic (RTK) GPS (Trimble Inc., Sunnyvale, CA, USA).
2. **Optimize alignment:** The internal and external camera positions were determined without GPS information, resulting in possible estimation errors. Georeferencing enables final optimization of the camera parameters and realignment of the sparse point cloud in this second step. GCPs permitted the highest accuracy for image alignment, in comparison to image GPS data (Agisoft 2018).
3. **Build dense cloud:** The coarse point cloud is used in the third step to condense the point cloud with depth information, based on the multiview stereo process. Resolution of the dense cloud is influenced by selected quality parameters. With high-quality settings, this is the most time-consuming and computationally demanding part of the photogrammetric process.
4. **Build DEM:** Based on the dense cloud, the digital elevation model (DEM) is calculated, which represents a surface model in the form of a grid of height values.

---

<sup>4</sup> Parts of this section have previously been published by Wilke et al. (2019) & Wilke et al. (2021)

5. Build orthomosaic: A geometrically corrected orthomosaic is generated based on the DEM, the reconstructed model and the surface elements of individual images. The blending mode is defined as mosaic. Therefore, areas in the orthomosaic try to be reconstructed based on the central area within recorded images.

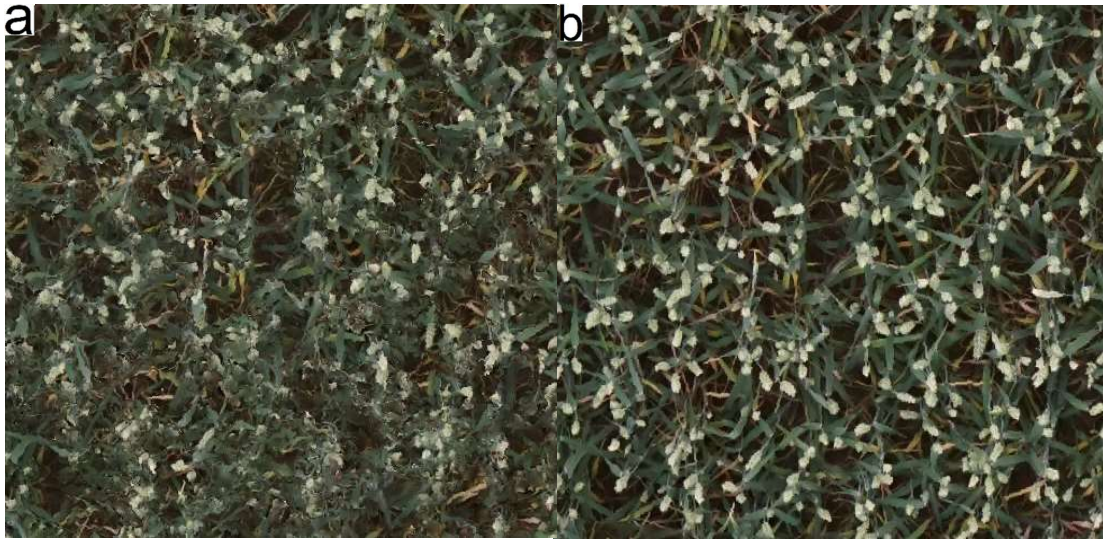
The processing settings were selected depending on the necessary source for trait assessment (e.g., lodging or plant density). For the 3D point cloud, as the main product, high-quality processing settings for image alignment and dense cloud generation were used (Table 3). This led to a point cloud with an adequate resolution to precisely extract canopy height in this thesis (Section 3). Using higher quality settings only increases the processing time without any benefit, while lower quality settings do not yield the required resolution for height extractions.

**Table 3** Ideal processing settings in Agisoft Professional depending on the target product (3D point cloud, orthomosaic) for trait extraction.

Step	Job type	Job type specifications	Processing settings	
			Product: 3D point cloud	Product: orthomosaic
1	Align photos	Accuracy	High	Low
2	Optimize alignment	Parameters	f, b1, b2, cx, cy, k1–k4, p1, p2	f, b1, b2, cx, cy, k1–k4, p1, p2
3	Build dense cloud	Quality	High	Lowest
		Filtering mode	Aggressive	Aggressive
4	Build DEM	Source data		Dense cloud
		Surface	/	DEM
5	Build orthomosaic	Surface		Mosaic
		Blending mode		Mosaic

In order to generate a high-quality orthomosaic, lower quality processing settings for image alignment and dense cloud generation were mandatory. The canopy structure, for instance, in the growth stage of spike development is very detailed and prone to possible errors in the reconstruction. With high-quality processing settings for image alignment and dense cloud generation, the number of matching features found within the SfM process also increases (Aasen et al. 2018), and the dense cloud naturally has a much higher resolution.

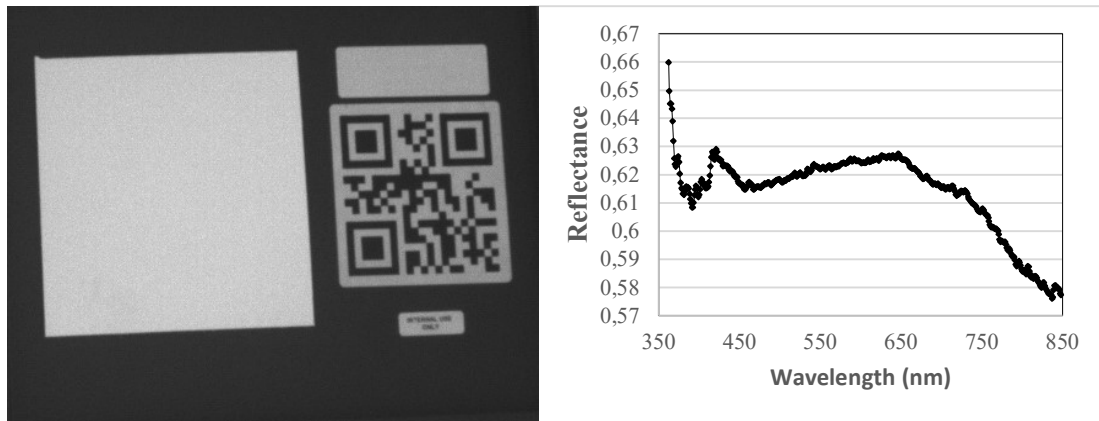
As a consequence, reconstruction of an area in the orthomosaic is likely to involve images with identical spikes in different positions, resulting in blurred and noisy areas (cf. Figure 7a.). This kind of issue can be avoided by choosing low-quality processing settings for image alignment and dense cloud generation, as illustrated in Figure 7b.



**Figure 7** Appearance of a similar area in a reconstructed orthomosaic influenced by different processing settings for image alignment and dense cloud generation: **(a)** high-quality processing settings resulting in blurry areas due to an increased number of matching features found for the SfM process; **(b)** low-quality processing settings for a high quality orthomosaic.

### 2.3.5 Calculating reflectance

The multispectral sensor records the digital numbers (DNs) of each spectral band. In order to calculate comparable physical units of reflectance, the aforementioned irradiance acquired by the DLS and the reference panel were used in this study. MicaSense provides an accurate radiometrical characterization of the panel (Figure 8), avoiding the need for own characterization systems like an expensive spectrometer.



**Figure 8** Reference panel provided by the camera manufacturer, radiometrically calibrated between 350 nm and 850 nm to calculate comparable physical units of reflectance for the multispectral sensor.

The data processing and calibration process is simplified and integrated in photogrammetric software like Agisoft (Agisoft 2019). Moreover, Agisoft automatically corrected images for exposure time, lens vignette effect, sensor gain, and black level values. The calibration process has been applied and evaluated successfully in different studies (Mamaghani et al. 2018; Su et al. 2019; Taddia et al. 2020). Finally, this setup enabled radiometric calibration of image data for a variation of irradiance during data acquisition, e.g., for cloudy illumination conditions.

### 3. Study I: Quantifying lodging percentage and lodging severity using a UAV-based canopy height model combined with an objective threshold approach

Norman Wilke <sup>1\*</sup>, Bastian Siegmann <sup>1</sup>, Lasse Klingbeil <sup>2</sup>, Andreas Burkart <sup>3</sup>, Thorsten Kraska <sup>4</sup>, Onno Muller <sup>1</sup>, Anna van Doorn <sup>1</sup>, Sascha Heinemann <sup>1</sup> and Uwe Rascher <sup>1</sup>

<sup>1</sup> Institute of Bio- and Geosciences, Plant Sciences (IBG-2), Forschungszentrum Jülich GmbH, 52428 Jülich, Germany

<sup>2</sup> Department of Geodesy, University of Bonn, 53115 Bonn, Germany;

<sup>3</sup> JB Hyperspectral Devices UG, 40225 Düsseldorf, Germany;

<sup>4</sup> Field Lab Campus Klein, Altendorf, University of Bonn, 53359 Rheinbach, Germany;

\* Correspondence: n.wilke@fz-juelich.de; Tel.: +49 2461 61-1733

Status: Published in Remote Sens. 2019, 11, 515.

<https://doi.org/10.3390/rs11050515>

**Author Contributions:** Conceptualization: Norman Wilke, Bastian Siegmann, Lasse Klingbeil, Andreas Burkart, Onno Muller, and Uwe Rascher; Designed the experiment: Thorsten Kraska, Onno Muller; Data curation, Norman Wilke, Andreas Burkart, and Sascha Heinemann; Formal analysis: Norman Wilke, Bastian Siegmann, and Andreas Burkart; Funding acquisition: Onno Muller; Investigation, Norman Wilke; Methodology: Norman Wilke, Bastian Siegmann, Lasse Klingbeil, Andreas Burkart, and Uwe Rascher; Project administration: Thorsten Kraska and Onno Muller; Supervision: Bastian Siegmann, Lasse Klingbeil, Andreas Burkart, Onno Muller, and Uwe Rascher; Validation: Norman Wilke and Sascha Heinemann; Visualization: Norman Wilke; Writing original draft, Norman Wilke and Bastian Siegmann; Writing review and editing: Norman Wilke, Bastian Siegmann, Lasse Klingbeil, Thorsten Kraska, Anna Van Doorn, and Uwe Rascher.

Original manuscript is embedded in dissertation format and redundancies were removed

#### Abstract

UAVs open new opportunities in precision agriculture and phenotyping because of their flexibility and low cost. In this study, the potential of UAV imagery was evaluated to quantify lodging percentage and lodging severity of barley using SfM techniques. Traditionally, lodging quantification is based on time-consuming manual field observations. Our UAV-based approach makes use of a quantitative threshold to determine lodging percentage in a first step. The derived lodging estimates showed a very high correlation to reference data ( $R^2 = 0.96$ , root mean square error (RMSE) = 7.66 %) when applied to breeding trials, which could also be confirmed under realistic farming conditions. As a second step, an approach was developed that allows the assessment of lodging severity,



information that is important to estimate yield impairment, which also takes the intensity of lodging events into account. Both parameters were tested on three ground sampling distances. The lowest spatial resolution acquired from the highest flight altitude (100 m) still led to high accuracy, which increases the practicability of the method for large areas. Our new lodging assessment procedure can be used for insurance applications, precision farming, and selecting for genetic lines with greater lodging resistance in breeding research.

### **3.1 Introduction**

The use of UAVs allow acquisition of image data with high spatial resolution, which is a basic requirement for deriving detailed 3D canopy structure of crops using feature matching and SfM techniques (Turner et al. 2012; Dandois and Ellis 2013; Colomina and Molina 2014). To derive the canopy height (CH) from the canopy structure, a nonvegetated ground model is needed. This ground model represents the topsoil surface, which is usually acquired by a UAV overflight shortly after sowing or harvest, when agricultural fields are free of vegetation (Bendig et al. 2013; Chu et al. 2017). This kind of information can also be provided by digital terrain models (DTMs) derived from LiDAR data. Different studies have already examined the quality and accuracy of LiDAR DTMs (Li 1988; Kraus et al. 2004; Podobnikar 2009; Hirt 2017). The potential of this data source as an alternative ground model for CH retrieval, however, has not yet been evaluated.

UAV CH has been investigated in several studies in recent years (Anthony et al. 2014; De Souza et al. 2017; Stanton et al. 2017). Noninvasive measurement of CH, multitemporal growth curves (Holman et al. 2016), and biomass development (Bendig et al. 2014) at different phenological stages is a major requirement for precision agriculture. Crop vitality, fertilizer status, soil quality, and water availability influence the aforementioned plant traits and are used to optimize agricultural management practices or yield predictions (Shanahan et al. 2001; Hansen and Schjoerring 2003; Adamchuk et al. 2010; Mulla 2013; Bareth et al. 2015). Compared to plant height measured with a ruler in the field at a specific location, the UAV-based approach provides information on the spatial height distribution of a continuous canopy (Aasen et al. 2015). Thus, it contains height information of numerous single plants, while reference measurements in the field only represent height information of single plants covering a very limited area.

Furthermore, CH is very well suited to quantify lodged areas. The term “lodging” is defined as permanent displacement of a plant from the upright position (Berry and Spink 2012; Rajapaksa et al. 2018). It is a major problem in cereal crops, because it leads to qualitative and quantitative yield losses up to 45 % (Weibel and Pendleton 1964; Pinthus 1974; Berry and Spink 2012; Peng et al. 2014). Yield loss is strongly affected by the lodging severity and the development stage at which it occurs (Fischer and Stapper 1987; Briggs 1990; Berry et al. 2004). Extreme weather conditions (wind, hail, heavy rain) and other environmental factors, such as excessive nitrogen supply, pests, diseases, and high plant density, can cause lodging before harvesting (Zhang et al. 2017; Dahiya et al. 2018). Thus,

breeding programs aim to select for genetic lines of cereal crops with greater lodging resistance (Pinthus 1974). Due to the impacts of climate change, with an increasing amount of extreme weather conditions, lodging is still a major limiting factor in yield impairment, resulting in an increasing need for a cost-efficient and accurate approach to quantify lodging. The use of UAVs can replace laborious and subjective ground data collection and enable spatial assessment of lodging by an automated system. Susko et al. (2018) tried to assess crop lodging with a field camera track system. Liu et al. (2018) used visible and thermal infrared images derived from UAV for rice lodging estimation. Texture information and canopy structure have also been used to assess lodging in rice (Yang et al. 2017b). Murakami et al. (2012) quantified lodging in buckwheat using the 3D canopy structure. In general, the area of lodging can be determined by using a predefined threshold at which CH lodging occurs. However, the thresholds applied in different studies (Chapman et al. 2014; Yang et al. 2017b) were defined by subjective inspections rather than by mathematical approaches.

The lodging percentage parameter identified with the single threshold approach only enables determining the presence or absence of lodging. In recent years, several studies showed that yield impairment is additionally affected by lodging severity (Fischer and Stapper 1987; Berry 1998; Berry and Spink 2012; Murakami et al. 2012). The term “lodging severity” is defined as the angle of the permanent displacement of plants from their upright position (Berry 1998; Berry and Spink 2012). Ground data based on visual lodging scores are generally insufficient in accuracy, efficiency, and objectivity (Simko and Piepho 2011; Murakami et al. 2012). So far, only Chu et al. (2017) have assessed the lodging severity of corn by quantifying the number of lodge plants. Due to the typical plant structure and plant density of corn, this approach cannot be applied to cereal crops.

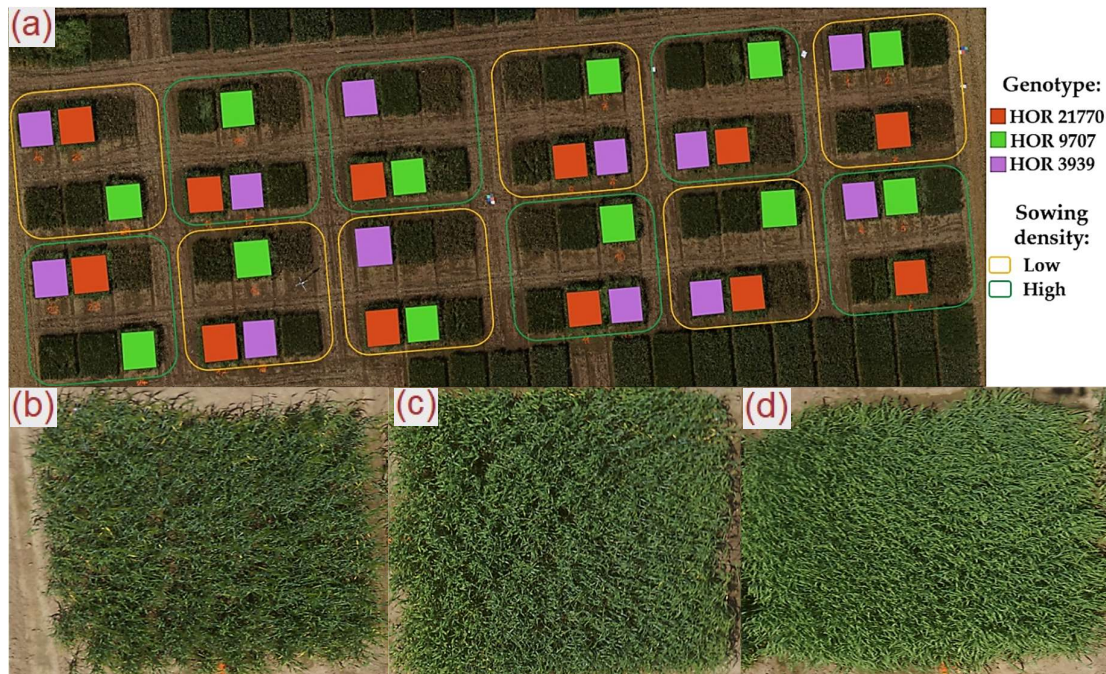
The major aims of this study are to use canopy structure derived from RGB image data to determine the factors that affect canopy height and to develop new methodologies for the assessment of lodging percentage and severity. The primary objectives are: (1) to compare plant traits derived from UAV- and DTM-based ground models to evaluate whether the DTM can serve as an alternative data source; (2) to investigate the comparability between ground-measured plant heights in the field with UAV-derived CHs, taking factors like canopy structure and plant density into account; (3) to develop a mathematical approach for the assessment of lodging percentage, avoiding adjusted thresholds and subjective decisions; and (4) to assess lodging severity with a novel approach based on canopy height variations.



## 3.2 Materials and methods

### 3.2.1 Study area

Within the research campus at the field lab on the Campus Klein-Altendorf of the University of Bonn, 2 experimental sites were investigated. The first experimental site was sown on 09 April 2016 consisted of several small breeder plots, each  $2.62 \times 3$  m in size. This experiment was the test bed for developing a method that allows canopy height and lodging assessment. The layout included 3 spring barley (*Hordeum vulgare*) cultivars (HOR 21770, HOR 9707, HOR 3939) and 2 different sowing densities with 6 repetitions (Figure 9a). The high density ( $300 \text{ seeds m}^{-2}$ ) reflected the common sowing density in Germany. The lower density consisted of  $150 \text{ seeds m}^{-2}$ . The selected barley cultivars varied in canopy characteristics (cf. Figure 9).

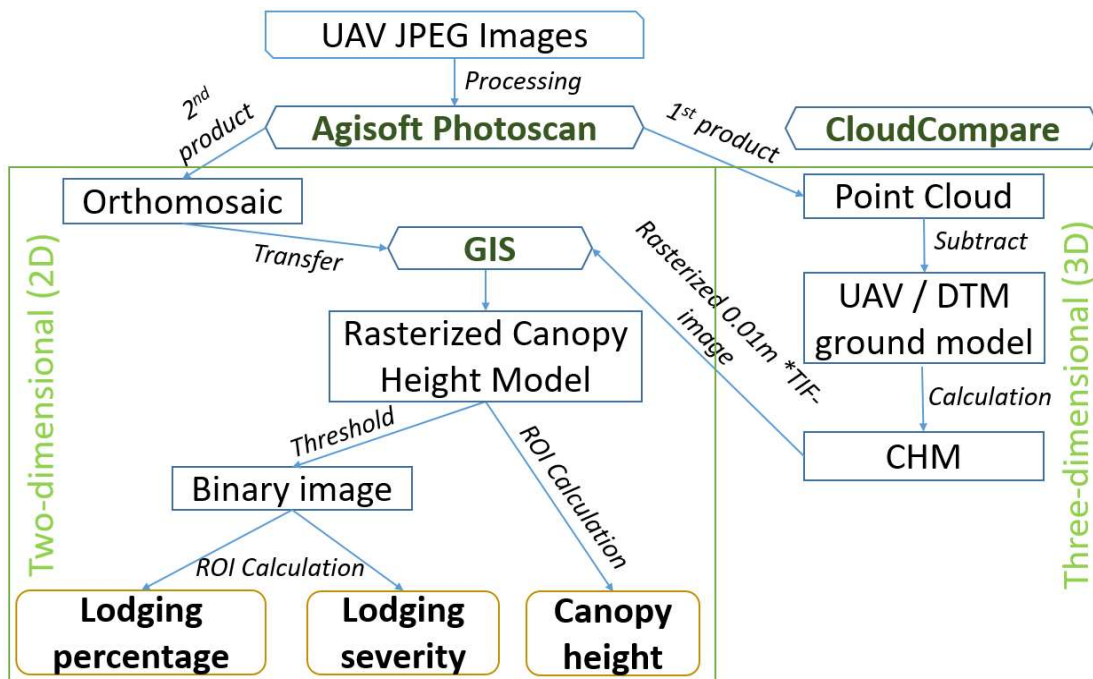


**Figure 9** (a) Experimental site 1, consisting of small breeder plots with three spring barley cultivars and two sowing densities. Representative images of three barley varieties 61 days after sowing (DAS); (b) genotype HOR 3939, characterized by sparser vegetation cover and small soil gaps; (c) HOR 9707, characterized by a closed canopy with fewer gaps and (d) HOR 21770, characterized by a closed and dense canopy without any gaps.

The second site was a normal production field with a size of 1.5 hectares (Figure 16a), where winter barley was sown on 28 September 2017. The variety “Lomerit” was sown with a commonly used density of  $300 \text{ seeds m}^{-2}$ . The lodging assessment approach developed at the first experimental site was applied to test its robustness and usability on this conventionally treated field. The experiments were fertilized with 170 kg nitrogen per hectare and growth inhibitors were applied. The seeds were sown in rows with 10.9 cm row distance.

### 3.2.2 Data acquisition, processing, and canopy height model generation

The Falcon-8 octocopter and the Sony Alpha 6000 RGB wide-angle camera with a 30 mm fixed focal length lens were used for image acquisition, resulting in a GSD of 0.54 cm at a flight altitude of 35 m above ground level (AGL). Agisoft PhotoScan (Version 1.4.1) was used to recreate the 3D point cloud (Westoby et al. 2012) as first product and the two-dimensional (2D) orthomosaic (Gómez-Candón et al. 2014) as second product. The evaluation of the 3D point cloud with the necessary processing settings (Table 3) was done in CloudCompare (Version 2.9.1) (Danielgm 2016). The 3D point cloud contains height information above the ellipsoid. To derive the canopy height model (CHM), the 3D point cloud must be subtracted from an underlying ground model (cf. Figure 10).



**Figure 10** Overall flowchart of software products used in this study for data processing and evaluation of the spatial assessment of canopy height, lodging percentage, and lodging severity. GIS: geographic information system; DTM: digital terrain model; CHM: canopy height model; ROI: region of interest.

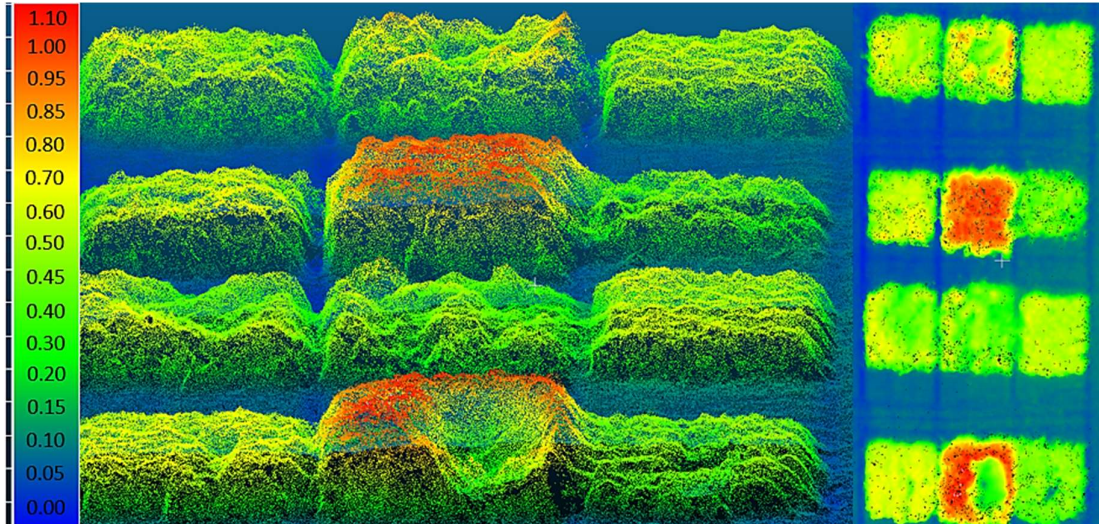
Two methods were investigated in this study to derive the ground model:

- 1) Ground model determination based on a UAV overflight shortly after sowing (UAV-based ground model), and
- 2) Ground model determination based on a DTM provided by state authorities (DTM-based ground model).

Method 1 requires an additional overflight in the time frame shortly after sowing until seedling emergence with no visible vegetation. In contrast, method 2 is based on the airborne

LiDAR dataset that provides the topsoil surface with a GSD of 1 m. In Germany, DTMs are freely accessible and updated at least every 6 years. The DTM for the study was acquired in 2015.

The height difference between the ground and the canopy model finally enables the assessment of canopy height (CH), as illustrated in Figure 11.



**Figure 11** Visualization of CHM (m) for selected barley plots in side view (left) and nadir top view (right).

Rasterization of the 3D CHM resulted in a 2D \*TIF image with high spatial resolution of 0.01 m (cf. Figure 10), from which the maximal height value for each grid cell was exported and used for further calculation.

### 3.2.3 UAV canopy height assessment and validation

The analysis of the rasterized CHM based on images with 0.85 cm GSD was done in QGIS (version 3.2.3) open source GIS software, using layers with different levels (cf. Figure 10). Shape files with a size of  $2.4 \times 2.8$  m were created for each barley plot of the experimental site and used to calculate the median CH. The size of the shape files was chosen slightly smaller than the size of the plots to avoid border effects in the further processing.

Plant heights were also determined with a measuring ruler in the field to compare them to the UAV- derived CH of experimental site 1. The median plant height was calculated based on 6 height measurements within each plot, whereby the highest point of the plants was taken. In total, 4 repetitions per sowing density and genotype were recorded 61 days after sowing (DAS).

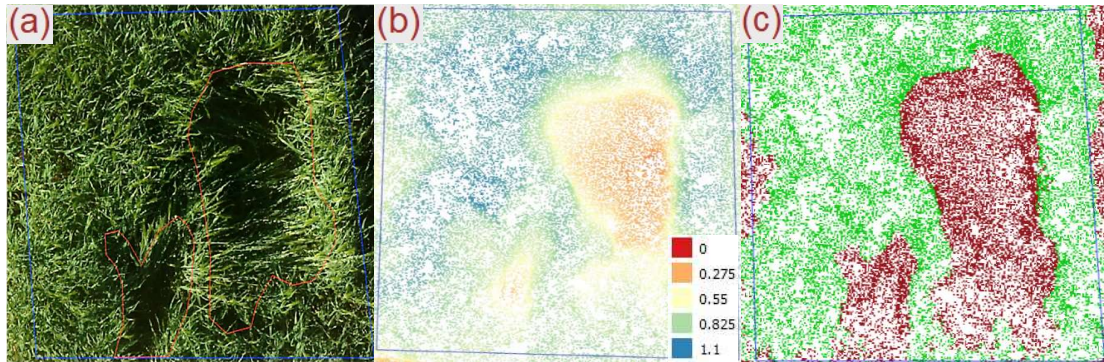


### 3.2.4 Lodging assessment and validation

The rasterized CHMs and the shape files ( $2.4 \times 2.8$  m) of the plots were also used for the lodging assessment (Figure 10). Two parameters were calculated in this study: (1) lodging percentage, which represents the area affected by lodging, and (2) lodging severity, which describes how strongly the canopy is affected by lodging based on the canopy height variation, hereinafter specified in detail.

#### 3.2.4.1 Experimental site 1: Breeding trials

The first step in the process of quantifying lodging was to determine an objective threshold. First, the maximum canopy heights (MAXCH) of each genotype were calculated based on images with 0.85 cm GSD. In the second step, the average MAXCH of all repetitions was used to minimize the risk of including outliers as maximum values. To determine the UAV lodging percentage, 4 lodging percentage thresholds (LPTs) related to the MAXCH were calculated: 80 % (80LPT), 70 % (70LPT), 60 % (60LPT), and 50 % (50LPT) of the MAXCH. For example, 70LPT of a MAXCH of 1 m is 0.7 m. Based on the different LPTs, the lodging percentage could be determined by a simple query (rasterized CHM < LPT), resulting in a binary image showing areas affected and not affected by lodging (Figure 12c).



**Figure 12** (a) High-resolution RGB orthomosaic used for manual validation of areas affected by lodging; (b) Calculated CHM (m) in the region of interest and (c) Binary image calculated based on 70 % lodging percentage threshold (70LPT) with areas affected (dark red) and not affected (green) by lodging.

For the lodging severity assessment a mixture of four thresholds (80LPT, 70LPT, 60LPT, 50 LPT) related to the MAXCH were used. Based on these thresholds, first the average lodging severity (ALS) and second the weighted average lodging severity (WALS) were determined according to Equations 1 and 2. In comparison to ALS, the WALS parameter additionally rates areas affected by lodging, differentiated to consider also the yield impairment. The applied weighting factors were chosen based on expertise, adjustments and the expected yield impairment of the LPTs. Sections with CH lower than 50 % (50LPT) of the MAXCH were weighted twice as much as those with CH lower than 80 % (80LPT) of the MAXCH (Equation 2). The difference between adjacent LPT factors within the WALS

calculation was 0.25, so that the value range for both lodging severity parameters varied between 0 and 100 %.

$$ALS = \frac{80LPT + 70LPT + 60LPT + 50LPT}{4} \quad (1)$$

$$WALS = \frac{(0.625 * 80LPT) + (0.875 * 70LPT) + (1.125 * 60LPT) + (1.375 * 50LPT)}{4} \quad (2)$$

To validate the accuracy of the UAV lodging percentage, affected areas of each barley plot were manually determined in a high-resolution orthomosaic (GSD = 0.23 cm). Due to the high spatial resolution, the lodged areas could be easily identified, resulting in a precise lodging percentage determination (Figure 12a). These reference data do not consider information on lodging severity and only provide a differentiation between the presence or absence of lodging.

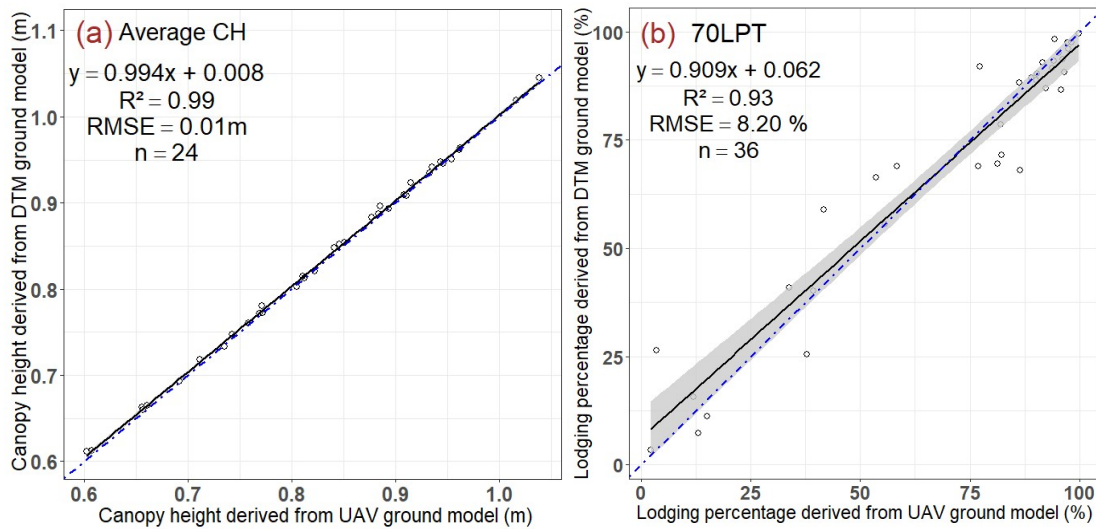
#### **3.2.4.2 Experimental site 2: Farmer field**

For the case study, the above described method was applied to a conventional production field. Due to the lack of repetition in a classical farmer field, the 90th percentile of canopy height distribution was used as MAXCH to minimize the risk of including an outlier as a maximum value. To evaluate the influence of the spatial resolution on the assessment of lodging percentage and severity, 3 datasets with different GSDs were acquired by adjusting flight altitude: 0.54 cm (35 m AGL), 1.09 cm (70 m AGL), and 1.57 cm (100 m AGL). Data acquisition took place 258 DAS. For validation purposes, the orthomosaic with 0.54 cm GSD was used to manually determine areas affected by lodging.

### 3.3 Results and analysis

#### 3.3.1 Comparison of plant traits derived from UAV- and DTM-based ground models

As aforementioned (cf. Section 3.2.2), the 3D point cloud has to be subtracted from a ground model to derive the CHM. The UAV- and DTM-based ground models were used in the first experiment (breeding trials) to determine the average CH (61 DAS) and lodging percentage parameter (75 DAS) for each barley plot. The CHs derived from both ground models showed a high level of agreement ( $R^2$  of 0.99) and provided almost the same results (Figure 13a).



**Figure 13** Scatter plots from UAV- and DTM-based ground models of experimental site 1 of **(a)** average canopy height 61 DAS and **(b)** lodging percentage (70LPT) 75 DAS. Black line represents regression line with 95 % confidence interval; blue line represents 1:1 line.

In comparison to CH, the lodging percentage parameter determined for both ground models using 70LPT resulted in a slightly lower  $R^2$  of 0.93 (Figure 13b). Small CH differences can influence the determined lodging percentage, especially for plots that are less affected by lodging. Therefore, a higher residual deviation can be observed for lower values by considering the 1:1 line (Figure 13b). In contrast, values higher than 50 % showed a better fit with the 1:1 line. Although the correlation of the parameter lodging percentage was slightly lower compared to the UAV CH, it still showed a high level of accuracy.

For experimental site 2, both ground models were investigated based on the parameter lodging percentage using 70LPT (Table 4). Only a small difference of 3.25 % between both ground models could be observed. Compared to the reference data, the UAV ground model showed a smaller deviation than the DTM ground model. Due to the slightly higher accuracy, the UAV ground model was used in the further course of the study.

**Table 4** Assessment of UAV lodging percentage (70LPT) derived from UAV- and DTM-based ground models in comparison to reference data for experimental site 2 258 DAS. LPT: lodging percentage threshold; DTM: Digital Terrain Model; DAS: Days after sowing.

Lodging percentage (%)		
UAV-based ground model	DTM-based ground model	Reference data
71.81	75.06	70.27

### 3.3.2 UAV canopy height assessment and validation

Comparing the UAV-derived CHs of the breeder trial with corresponding plant heights measured directly in the field, clear deviations in accuracy could be detected, depending on genotype and sowing density. While genotype HOR 3939, with 0.29 m in low density and 0.18 m in high density, showed greater differences between UAV CH and reference measurements, genotypes HOR 9707 and HOR 21770 showed only slight differences (Table 5).

**Table 5** Comparison of UAV CH and reference measurement 61 DAS (n = 24). SD: standard deviation.

Genotype	Sowing density	Median and SD (m)		Discrepancy between reference measurements and UAV CH (m)
		Reference measurements	UAV CH	
HOR 3939	Low	0.96 ± 0.02	0.67 ± 0.08	(-) 0.29
HOR 9707		1.00 ± 0.04	0.92 ± 0.05	(-) 0.08
HOR 21770		0.93 ± 0.02	0.90 ± 0.05	(-) 0.03
HOR 3939	High	0.94 ± 0.05	0.76 ± 0.06	(-) 0.18
HOR 9707		1.02 ± 0.03	0.99 ± 0.04	(-) 0.03
HOR 21770		0.93 ± 0.01	0.92 ± 0.01	(-) 0.01

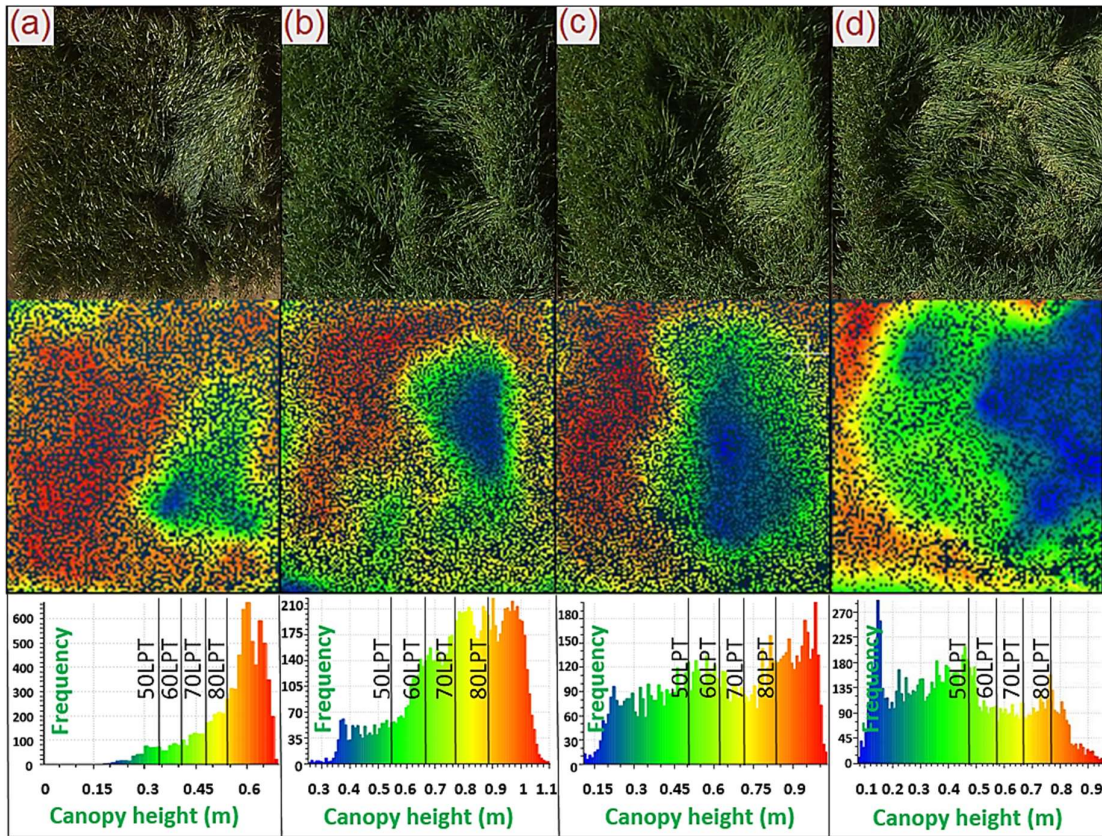
One reason for the more pronounced deviations observed for genotype HOR 3939 was that the canopy fractional cover was lower than that of the other two cultivars. This resulted in small areas inside the plot without vegetation cover where soil was visible (Figure 9b). Thus, not only the top layer of the canopy was acquired, but lower parts were also included in the computation of CHM. These nonvegetated areas had an influence on the determined median CH and caused underestimation of the UAV-based CH retrieval. In contrast, genotype HOR 21770 showed a dense and closed canopy without any gaps (Figure 9d). Thus, only the top canopy layer was considered in CHM generation, consequently leading to a small deviation of 0.03 m in the low sowing density and 0.01 m in the high sowing density compared to the reference measurements (Table 5). In general, the UAV CH of the high-density plots matched better with the plant heights measured in field with an average difference of 0.07 m in comparison to the lower-density plots with an average difference of 0.13 m. This is also due to the fact that the lower sowing density resulted in a less dense canopy with gaps and consequently influenced the UAV CH assessment.



### 3.3.3 Lodging assessment and validation

#### 3.3.3.1 Experimental site 1: Breeding trials

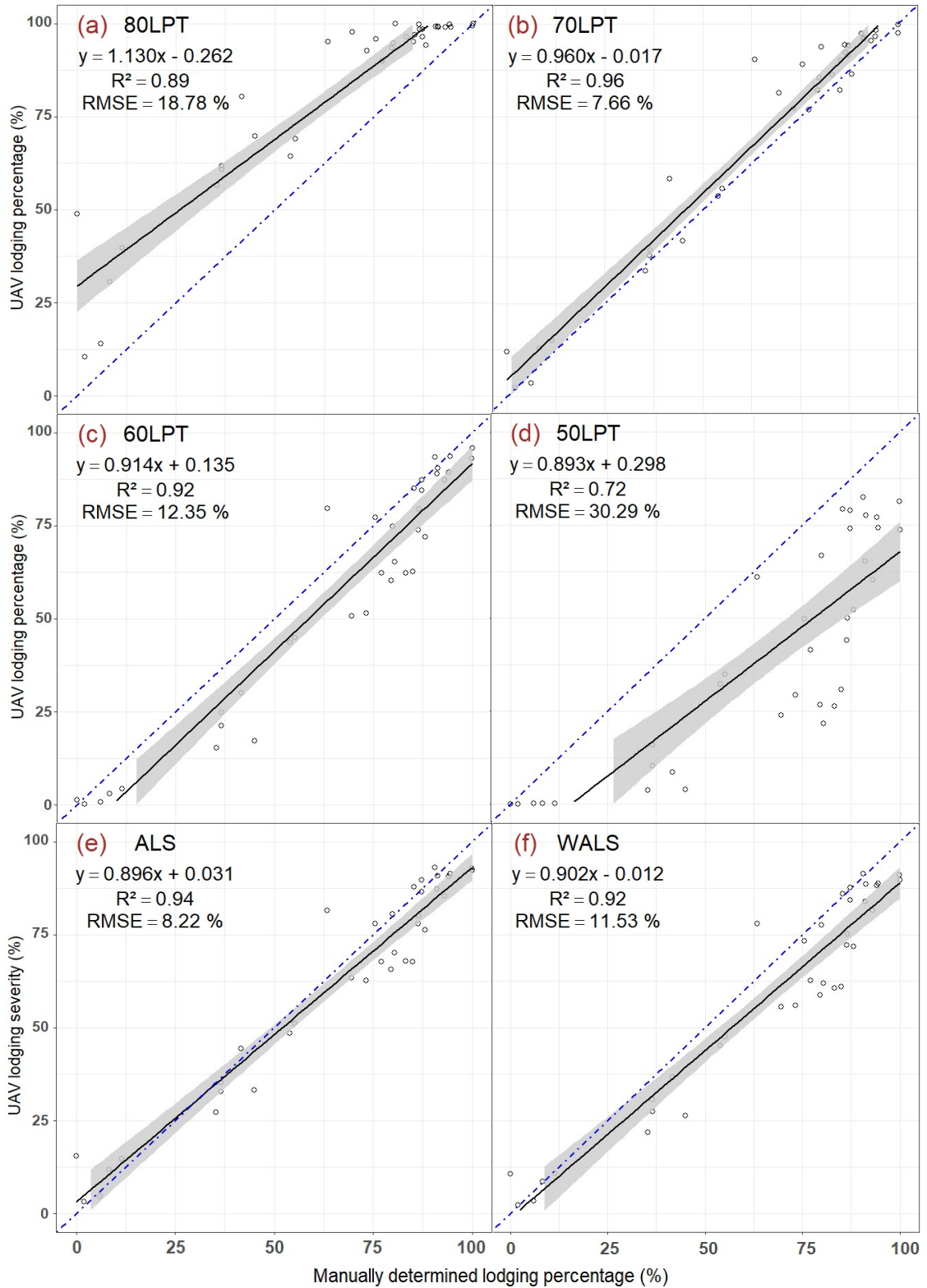
Permanent displacement of a plant from the upright position can be affected by different lodging severity, as illustrated in Figure 14. The first example shows a slightly affected canopy where only low-yield loss can be expected (Figure 14a). The same applies for the plot in Figure 14b, where just a few small areas are strongly affected by lodging. Additionally, Figure 14c clearly illustrates the capability of the approach to detect lateral lodging. In contrast to the other examples, the canopy in Figure 14d is heavily affected by lodging, so that parts of the plot lay completely on the ground.



**Figure 14** RGB images of barley plots showing different types of lodging (top) and corresponding lodging severity derived from the CHM (middle), as well as canopy height distributions with visualization of different lodging percentage thresholds (LPTs) (bottom).

In order to identify the most suitable threshold for lodging percentage assessment, 4 LPTs (80LPT, 70LPT, 60LPT, 50LPT) were compared to the reference data. The UAV lodging percentage derived from 80LPT has a RMSE of 18.78 % (Figure 15a).





**Figure 15** Scatter plots of manually determined lodging percentage and calculated UAV-based lodging percentage for (a) 80LPT; (b) 70LPT; (c) 60LPT; (d) 50LPT as well as lodging severity parameters (e) ALS, and (f) WALS 75 DAS. Black line represents regression line with 95 % confidence interval; blue line represents 1:1 line (n = 36).

It became clear that the absolute height differences between MAXCH and 80LPT were small and varied between 0.13 m (HOR 3939, high sowing density) and 0.22 m (HOR 21770, low sowing density) (Table 6). The naturally occurring plant height variation was higher than the predefined lodging threshold, resulting in an overestimation of lodging (Figure 15a).

**Table 6** Overview of MAXCH, UAV lodging percentage for four LPTs (80 %, 70 %, 60 %, 50 %), ALS, WALs, and manually determined lodging percentage reference data for different sowing densities and genotypes 75 DAS (n = 36). MAXCH: maximum canopy height; LPT: lodging percentage threshold; Ref: Reference data; WALs: weighted average lodging severity; ALS: average lodging severity.

Genotype	Sowing density	MAX CH (m)	Lodging percentage (%)					Lodging severity (%)	
			80LPT	70LPT	60LPT	50LPT	Ref.	WALS	ALS
HOR 3939	Low	0.72	74.70	59.94	41.74	20.76	53.97	43.66	49.29
HOR 9707		0.79	84.90	70.54	54.35	34.48	70.54	55.84	61.07
HOR 21770		1.12	44.59	26.86	16.21	9.77	24.81	20.76	24.35
HOR 3939	High	0.66	94.52	86.90	73.00	50.10	77.27	71.53	76.13
HOR 9707		0.68	98.10	92.86	80.94	58.44	73.28	78.49	82.58
HOR 21770		1.03	92.45	85.75	78.30	69.37	80.90	79.07	81.47

The UAV lodging percentage derived from 70LPT took into account the naturally occurring plant height variation in the field and led to the highest correlation ( $R^2 = 0.96$ ) and the lowest RMSE (Figure 15b). The low amount of scattering indicated that 70LPT can be applied independently from the amount of lodged plants. No influence of the aforementioned differentiated canopy characteristics and canopy heights was observed. In comparison to the reference data, the UAV lodging percentage derived with 60LPT and 50LPT showed lower correlations (Figure 15c,d). Canopy areas affected by lodging were partly not identified when the lower CH thresholds were applied. Thus, the lodging percentage was underestimated, especially in strongly affected plots, where the CH threshold was more relevant.

In comparison to the single threshold approach normally used, a combination of different thresholds can provide additional information for the lodging percentage assessment. The average of all four thresholds used in the ALS calculation enabled estimation of the lodging percentage on a pixel basis with high accuracy ( $R^2 = 0.94$ , RMSE = 8.22 %) (Figure 15e). The WALs parameter, in contrast, provided slightly lower accuracy ( $R^2 = 0.92$ , RMSE = 11.53 %) (Figure 15f), because of the weighting procedure that was implemented within the parameter calculation (Equation 2).

The aforementioned lodging severity variation cannot be determined by applying a single threshold approach, because that only represents a binary distinction between lodged and nonlodged areas. Consequently, 70LPT, for example, cannot distinguish between slightly affected areas (Figure 14a,b), where a low-yield impairment can be expected, and heavily affected areas (Figure 14d), which necessarily cause yield loss.

As a second step, the WALS parameter took these CH variations into account by its inbuilt weighting procedure and therefore can be used as an indicator for yield impairment. The reference data displayed in Table 6 clearly show that the lower sowing density with an average of 50 % were less affected by lodging compared to plots with the higher sowing density averagely affected by 77 %. Moreover, 50LPT applied to the lower-density plots allowed detection of only 35 % of lodged area at maximum and 10 % at minimum. Contrarily, 70LPT determined a distinctly higher amount of 71 % lodge area at maximum and 27 % at minimum.

The applied weighting procedure within the WALS calculation based on the different thresholds was able to consider this lodging intensity variation and, compared to the lodging percentage derived from 70LPT, led to a difference of 16 % at maximum (Table 6). The disparity between the WALS and reference data was influenced by the fact that the determined reference data were only an indicator for the presence or absence of lodging and did not provide information on lodging severity. The plots with high sowing density showed distinctly larger areas heavily affected by lodging, with 69 % at maximum and 50 % at minimum for 50LPT. This higher lodging severity led to a stronger correlation between the WALS and determined reference data (Table 6). Nevertheless, the variations still present between the different LPTs resulted in a deviation of 15 % at maximum between 70LPT and WALS for the plots with high sowing density. The average difference considering all genotypes and sowing densities between both parameters was 12 % ( $SD \pm 4.2$ ).

Comparing ALS and WALS clarified that ALS without the weighting procedure slightly overestimated the lodging severity (Table 6). The difference between WALS and ALS was 6 % at maximum, the average difference was 4 % ( $SD \pm 1.1$ ).

### 3.3.3.2 Experimental site 2: Farmer field

The results for the developed lodging assessment procedure applied to an entire farmer field for different GSDs are summarized in Table 7. The 70LPT calculated for the highest spatial resolution (0.54 cm) showed the closest match to the reference data. Therefore, 70LPT again seemed to be most suitable to assess the lodging percentage.

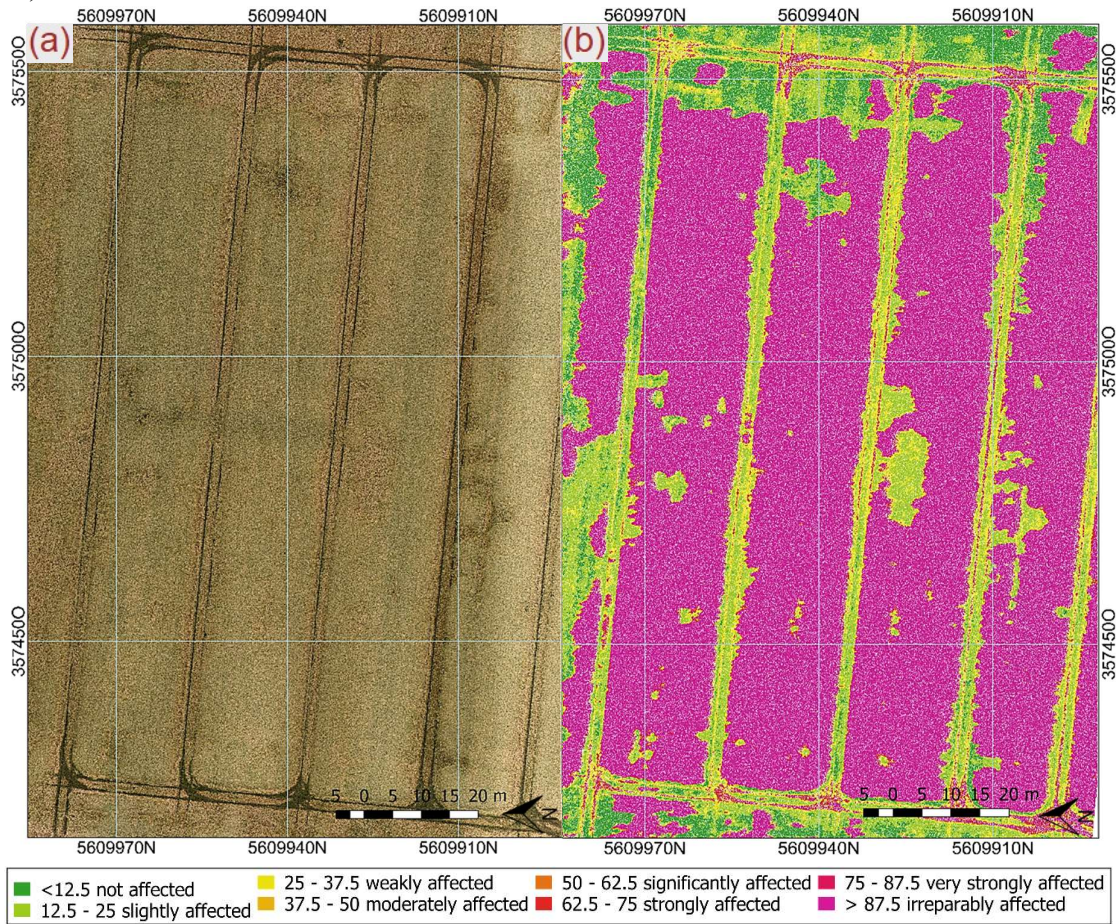
**Table 7** Comparison of lodging percentage reference data with UAV-based lodging percentage for four LPTs (80 %, 70 %, 60 %, 50 %) and UAV lodging severity for image datasets with different spatial resolutions.

GSD (cm)	Lodging percentage (%)				Lodging severity (%)		Reference data
	80LPT	70LPT	60LPT	50LPT	WALS	ALS	
0.54	88.83	71.81	66.69	64.75	70.61	73.02	70.27
1.09	89.79	78.04	68.11	64.36	72.38	75.08	
1.57	87.35	78.51	73.05	68.60	74.95	76.88	

The results for 70LPT determined for the image data with lower spatial resolutions (1.09 and 1.57 cm) led to slightly higher deviations at around 6 % compared to the reference data. However, the measured deviations were quite small. The general trend showed that the lodging percentage of all LPTs increased with decreasing GSDs. Therefore, the results for 60LPT correlated increasingly better with the reference data with lower spatial resolution. This was influenced by the fact that the lodging severity of the entire farmer field was relatively high (Figure 16) and the differences between LPTs were small.

The calculated lodging severity parameters showed exactly the same trend regarding the different spatial resolutions. The WALS parameter only increased by 4 % with decreasing GSDs. This underlines that the GSD has only a small impact on the parameter calculation. The WALS calculated for the dataset with the highest GSD (cf. Figure 16) provided a high level of agreement compared to the reference data. The small deviations of only 0.35 % for the reference data and 1.2 % for 70LPT were almost negligible and again were influenced by the high amount of plants strongly affected by lodging determined for 50LPT (cf. Table 7).

The ALS once more probably slightly overestimated the lodging severity because of the missing weighting procedure. The difference between the two parameters, however, was only 3 % at maximum because of the small variations between the different LPTs (cf. Table 7).



**Figure 16** Experimental site 2: (a) conventional farmer field 1.5 hectare in size and (b) result of WALs determination (0.54 cm GSD) 258 DAS.

### 3.4 Discussion

In the presented study, different methods for high-throughput field phenotyping based on UAV image data were developed. In the following, the results achieved for the investigations are discussed in detail: a comparison of UAV-based and DTM-based ground models, comparability between measured plant heights in the field and UAV-derived CHs, and a spatially precise assessment of lodging percentage and severity.

The quality assessment of DTMs derived from airborne LiDAR data was mostly based on statistical methods (Li 1988; Kraus et al. 2004; Podobnikar 2009). Comparing the LiDAR DTM in this study with a UAV-acquired ground model, the results showed that the LiDAR DTM can be used as an almost equivalent ground model to determine accurate CHMs. A direct comparison of the ground models independent from plant traits was not done, because the average height difference between the two ground models would consider irrelevant information. Agricultural activities such as deeper wheel tracks were more pronounced in the UAV ground model because of the distinctly higher spatial resolution in comparison to the LiDAR DTM. Areas like this were not relevant for the CH and lodging assessment and therefore can be neglected. The results in this study are independent from the year (2016 – Experimental site 1, 2017 – Experimental site 2) and location of data collection. Without extensive changes caused by human impact or agricultural crop rotation, it can be assumed that the renewed interval every six years provides sufficiently up-to-date information to obtain a precise ground model. However, a larger-scale experiment considering different locations and years needs to be conducted in the future to exploit the potential of DTM-based ground models in more detail. Nevertheless, the DTM has the potential to substitute for the mandatory UAV overflight and thus substantially reduce the effort in data collection. In this way, the time needed for data collection and data processing to determine lodging can be reduced by half. The included height information of the topsoil surface every square meter can be a benefit for large fields or hilly areas.

As already demonstrated in different studies (Bendig et al. 2014; Holman et al. 2016; Chu et al. 2017), the UAV-based CH can be highly correlated with manually determined plant heights collected with a measuring ruler in the field. This study additionally shows that the UAV-derived information on the spatial height distribution of the canopy was affected by the canopy structure, the sowing density, and correspondingly also the sowing heterogeneity. In contrast, manual plant height measurements only represent single points in the field associated with subjective decisions, in particular with high plot heterogeneity. For that reason, the UAV-derived CH should be considered as an autonomous trait with a different definition compared to the plant height measurements. A validation of the UAV CH using the plant height measurements was correspondingly only possible to a limited extent. The UAV assessment enabled representative information on plots with high heterogeneity that can be measured manually only with great effort. Especially for breeders, the CH determination enabled simplifying complex crop and plant surfaces on an objective scale to estimate genetic effects. Furthermore, canopy homogeneity can be investigated for



entire farmer fields because it carries information on the heterogeneity of the underlying soil and the factors influencing crop growth.

The developed approach for lodging percentage assessment provided very high accuracy in breeding trials ( $R^2 = 0.96$ , RMSE = 7.66 %) and led to a slight overestimation of 2 % when applied to a classical farmer field. Compared to other approaches where thresholds are normally chosen based on subjective decisions without a mathematical approach and reference data for validation (Chapman et al. 2014; Yang et al. 2017b), 70LPT enabled the precise detection of lodged areas within the canopy and took into account the naturally occurring plant height variations in the field. Furthermore, the implemented method for detection of an objective threshold considered the aforementioned factors influencing the CHM. The results showed that the developed method is well suited for barley genotypes with differentiated canopy structures and therefore has the potential to be applied to other cereal crops, such as wheat.

In the process of determining lodging percentage in rice using structure, texture, and thermal information derived from UAV images, Liu et al. (2018) and Yang et al. (2017b) obtained high  $R^2$  values greater than 0.9. The accuracy of the lodging percentage determined from textural information, however, is strongly dependent on a trained support vector machine (SVM) and the dataset used. Changing illumination conditions during the flight, general illuminance, sun angle, shadow effects, plant development stages, and color variance between genotypes and species can influence the method. For that reason, the approach is not transferable without having to adapt the SVM to other datasets. The quantification of lodging from thermal images is also very challenging, because external factors such as small changes in wind speed and cloud cover can strongly influence the derived canopy surface temperatures (Jackson et al. 1983; Chapman et al. 2014). The temperature difference between lodged and nonlodged plants was quite low, which can only be determined from image data recorded by precisely calibrated camera systems combined with accurate processing from raw data to final products.

The lodging assessment based on RGB images to derive the relevant CH presented in this study is almost independent from abiotic and external factors. Just a consumer RGB camera is needed, without the demand for calibration. In general, only large canopy height variations within a field can cause problems. In this extreme case, lower grown plants would be labeled as lodged plants. This issue, however, can be considered in the workflow by applying differentiated MAXCH values in areas with strong CH variations caused, for example, by different soil or nutrition conditions.

An advantage of the newly developed WALS and ALS lodging severity parameters is that they additionally take CH variance into account and enable quantifying of yield impairment caused by lodging. Several studies only considered the presence or absence of lodging, and different lodging severities as illustrated in Figure 14, were treated equally (Chapman et al. 2014; Yang et al. 2017b; Liu et al. 2018). Already Fischer and Stapper (1987) and Berry and Spink (2012) showed that the yield potential was influenced by the intensity of the permanent displacement of crops from their upright position. Additionally,

Murakami et al. (2012) investigated the usability of UAV data to assess lodging and showed that the yield was stronger impaired by higher lodging scores and low average CHs. Taking this into account, the novel WALs parameter was designed to consider the influence of lodging on yield. For the general lodging severity assessment the ALS parameter is adequate. However, to quantify the yield impairment caused by lodging, the LTP has to be weighted (WALS) to improve the prediction accuracy. In future studies, the factors applied to lower LPTs (Equation 2) should be ideally weighted more strongly and compared to yield data to investigate the potential in more detail.

The average difference of 12 % between the lodging percentages derived from 70LPT and WALs of experimental site 1 illustrate the need to differentiate between the lodging percentage and lodging severity. Even though plots with high sowing density were partly strongly influenced by lodging, there was still variation between the different LPTs, resulting in a reasonable deviation between the two parameters (cf. Table 6). This discrepancy between 70LPT and WALs decreased with less divergent LPTs, as ascertained for the farmer field of experimental site 2 (cf. Table 7). The higher the deviation between LPTs, the higher the difference between 70LPT and WALs. For the lodging severity in general the ALS parameter is more objective without the weighting factors. However, for the yield impairment caused by lodging the LTP has to be weighted to improve the prediction.

The results showed finally, that detection of lodge areas was still possible with the lowest spatial resolution (1.6 cm GSD) from the highest flight altitude (100 m) without a substantial decline in accuracy. Nevertheless, for very high accuracy, it is recommended to use images with higher spatial resolution (0.5 cm GSD), otherwise small patches with differentiated lodging severity will hardly be detected and severity values will increase.

To summarize, the developed lodging assessment approach can be used for insurance applications, precision farming, and breeding research. Besides selecting for genetic lines with higher lodging resistance, the different lodging severities and consequently yield impairments can be quantified as additional information. The approach additionally enabled determination of the recovery rate of crops. Navabi et al. (2006) demonstrated on over 140 different wheat genotypes that the extent of recovery capability varied among genotypes. Similar results were found by Briggs (1990) for barley.

The pixel-based lodging severity information based on the WALs parameter, illustrated in Figure 14, can be further used in precision farming to generate harvest maps and improve yield quality by avoiding areas in the harvest process that sprout again after heavy lodging events during the early stages of plant development.



### **3.5 Conclusions**

In the meantime, UAV technology is widely used because the data acquisition is relatively easy, timely, flexible, and cheap. The acquired data can provide timely, detailed information on the current status of plants, which is valuable for breeders, insurance companies, and farmers. Breeding trials are particularly difficult to monitor on a regularly basis within a reasonable time, resulting in an increasing need for faster selection of superior lines.

First steps were realized in this study to use an airborne LiDAR DTM provided by national authorities as an alternative ground model for CHM generation. The comparison of the DTM with the UAV ground model demonstrated that the DTM information can be used as a ground model and can help to reduce the effort in data collection and processing. Further investigations are needed to evaluate the robustness of ground models derived from LiDAR data under different conditions and in different locations.

The UAV-based CH assessment provides spatial information on the canopy height distribution and offers much more information compared to the classical plant height measurements of single spots in the field. The UAV-derived CH enables simplification of a complex crop surface with an objective scale to estimate genetic effects.

Moreover, the presented lodging assessment approach based on 3D canopy structure has many advantages over other methods, because it is more independent from external conditions, which increases its practicability. Furthermore, the method makes it possible to estimate yield impairment caused by lodging. Future studies need to be conducted to evaluate the accuracy in more detail. Finally, it was shown that areas affected by lodging could be detected with high accuracy even at the lowest spatial resolution (1.57 cm GSD). The higher the flight altitude, the shorter the flight time, the smaller the number of recorded images, and the shorter the processing time. Therefore, fixed-wing UAVs, normally operated at higher altitudes to cover large areas, can be used for lodging assessment. This also substantially increases the practicability of the developed method, especially for large agricultural fields.

## 4. Study II: Assessment of plant density for barley and wheat using UAV multispectral imagery for high-throughput field phenotyping

Norman Wilke <sup>1,\*</sup>, Bastian Siegmann <sup>1</sup>, Johannes A. Postma <sup>1</sup>, Onno Muller <sup>1</sup>, Vera Krieger <sup>1</sup>, Ralf Pude <sup>2,3</sup> and Uwe Rascher <sup>1</sup>

<sup>1</sup> Institute of Bio- and Geosciences, Plant Sciences (IBG-2), Forschungszentrum Jülich GmbH, 52428 Jülich, Germany

<sup>2</sup> Field Lab Campus Klein, Altendorf, University of Bonn, 53359 Rheinbach, Germany

<sup>3</sup> Institute for Crop Science and Resource Conservation, Renewable Resources, University of Bonn, 53115 Bonn, Germany

\* Correspondence: n.wilke@fz-juelich.de; Tel.: +49 2461 61-1733

Status: Published in Computers and Electronics in Agriculture. 2021, 198.  
<https://doi.org/10.1016/j.compag.2021.106380>

**Author Contributions:** Norman Wilke: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. Bastian Siegmann: Conceptualization, Formal analysis, Supervision, Visualization, Writing – original draft, Writing – review & editing. Johannes A. Postma: Formal analysis, Supervision, Writing – review & editing. Onno Muller: Writing – review & editing, Supervision, Funding acquisition. Vera Krieger: Writing – review & editing. Ralf Pude: Writing – review & editing, Supervision. Uwe Rascher: Writing – review & editing, Funding acquisition, Project administration, Supervision, Resources, Investigation.

Original manuscript is embedded in dissertation format and redundancies were removed

### Abstract

Cereal plant density is a relevant agronomic trait in agriculture and high-throughput phenotyping of plant density is important for the decision-making process in precision farming and breeding. It influences the water as well as the fertilization requirements, the intraspecific competition, and the occurrence of weeds or pathogens. Recent studies have determined plant density using machine-learning approaches and feature extraction. This requires spatially very highly resolved images (0.02 cm) because the accuracy distinctly decreased when images had lower resolution. In this study, we present an approach that uses the linear relationship between plant density manually counted in the field and fractional cover derived from a RGB and a multispectral camera equipped on an UAV. We assumed that at an early seedling stage fractional cover is closely related to the number of plants. Spring barley and spring wheat experiments, each with three genotypes and four different

sowing densities, were examined. The practicability and repeatability of the methodology were evaluated with an independent experiment consisting of 42 winter wheat genotypes. This experiment mainly differed for genotypes, sowing density and season.

The empirical regression models that make use of multispectral images having a GSD of 0.69 cm were able to determine plant density with a high prediction accuracy for barley and wheat ( $R^2 > 0.91$ , mean absolute error (MAE)  $< 28$  plants  $m^{-2}$ ). In addition, prediction accuracy only slightly declines for multispectral image data having 1.4 cm GSD or RGB image data having 0.6 cm GSD (MAE  $< 35$  plants  $m^{-2}$ ). BBCH stage 13 was identified as the ideal growth stage in which the plants were large enough to accurately determine fractional cover even from the lower resolution image data. Moreover, the developed empirical regression model transferred to an independent experimental field verifying its robustness across different conditions. The prediction accuracy of UAV estimated plant density showed an  $R^2$  value of 0.83 and an MAE of less than 21 plants  $m^{-2}$ . Furthermore, manual measurements of 11 randomly selected plots proved sufficient for a user-based training of the regression model ( $R^2 = 0.83$ , MAE  $< 23$  plants  $m^{-2}$ ).

The method and the use of UAV image data enable high-throughput phenotyping of cereal plant density with uncertainties of less than 10 %. The practicability, repeatability, and robustness of the developed approach were demonstrated in this study.

## 4.1 Introduction

The non-invasive assessment of plant traits with new sensors and autonomous small aircraft is becoming increasingly important in agriculture. The quantification of plant density under field conditions is also part of this ongoing development. Water and fertilization requirements are dependent on this parameter for agricultural management purposes. In addition, plant density has an impact on intraspecific competition and the occurrence of weeds or pathogens (Olsen et al. 2006; Olsen and Weiner 2007; Lawles et al. 2012; Liu et al. 2017a; Aich et al. 2018). A homogenous and ideal plant density is an important prerequisite for efficient crop production and potential yield (Joseph et al. 1985; Whaley et al. 2000; Valério et al. 2009; Liu, et al. 2017a; Jin et al. 2017a; Postma et al. 2020).

Plant density depends on sowing density and germination rate. Most yield prediction models typically consider sowing density as a trait rather than plant density. Although the germination rate is often corrected for in sowing density this only takes into account the genotypic variation in germination rate determined under optimal conditions and not the influence of abiotic (temperature, moisture, soil, nutrients, and frost), biotic (pests and diseases), and management factors (sowing variations, e.g. depth) (Lindstrom et al. 1976; Winter et al. 1988; Steiner et al. 1989; Kirby 1993; Murungu 2011; Rajala et al. 2011; Al-Mulla et al. 2014). Yield predictions can be improved by substituting estimated plant density for sowing density. However, such an implementation in modeling or decision-making is only feasible and realistic with a working high-throughput phenotyping approach (Furbank and Tester 2011; Araus and Cairns 2014). The current practice of representing the conditions

in a field or a plot is to count individual plants within several smaller subsample areas. This approach requires agricultural experts, is time-consuming, expensive, and not representative in the case of spatial variability within the field.

To replace these laborious ground data collection methods, the plant density has already been determined using different methods for maize (Shrestha and Steward 2003; Tang and Tian 2008), potatoes (Sankaran et al. 2017; Li et al. 2019) and sugar beet (Lottes et al. 2017). The planting structures of these cultivars with evenly spaced seedlings prevent an overlap among neighboring plants in early developmental stages and allow a relatively easy quantification of the germination rate. For cereals, however, mechanical seed drilling with a non-even seed distribution is the standard practice. Plants emerge very close to each other with narrow leaves overlapping among neighboring plants and single plants develop in multiple tillers further complicating individual plant identification especially at later stages. Therefore, the aforementioned methods developed for crops with a clear row structure cannot be reliably applied to cereal crops.

The plant density of cereals were also determined in previous studies from images recorded by RGB cameras mounted on an UAV, field bicycle or monopod (Jin et al. 2017b; Liu et al. 2017a; Liu et al. 2017b). Machine-learning techniques based on feature extraction were used to estimate plant density. High-resolution image data (0.02 cm) were required in these studies to enable a good prediction accuracy with feature extraction (Jin et al. 2017b; Liu et al. 2017a; Liu et al. 2017b). In this context, Jin et al. (2017b) have already demonstrated a distinctly decreasing accuracy using image data with a lower spatial resolution (0.10 cm). The prediction accuracy was also dependent on the sowing density with decreasing performance in the case of higher sowing density and probability of overlapping plants (Liu et al. 2017a; Liu et al. 2017b). In summary, the aforementioned approaches only permit a low throughput and limited adaptability to high sowing frequency.

In the present study, the plant density determination of cereals is based on fractional cover estimated from UAV multispectral image data. At the early seedling stage, plant density does not affect individual plant size. Only the area of overlapping leaves increases, especially with higher plant density. However, even at high plant densities, we expected that there would be a relatively little surface area overlapping as leaves are narrow. We assumed that a higher value of fractional cover at the early seedling stage also indicates a larger number of plants and a linear relation. Using this hypothesis, we trained empirical regression models for barley and wheat based on reference measurements acquired in the field and UAV derived fractional cover values. The procedure does not require the assignment of single leaves to particular seeds. This enabled the determination of plant density from spatial image data with potentially lower resolution leading to higher practicability of the methodology. Sankaran et al. (2015) had already made use of pixel values of a calculated vegetation index for germination assessment with a good correlation to ground truth observations. However, the visual ground truth rating of germination in their study with 10 % increments was only a rough estimation of plant density. Moreover, the development of

an empirical regression model and the transferability of the model to unknown data was not investigated.

The following study aims to answer the following questions, which can be divided into factors influencing the accuracy of the empirical regression model (i to v) and the evaluation of transferability of the method to an independent test site (v, vi): (i) What is the ideal plant developmental stage for data collection? (ii) What spatial resolution is necessary to achieve both high prediction accuracy and high throughput? (iii) What influence do plant characteristics (e.g., leaf arrangement, species, genotype) have on the methodology? (iv) Is an RGB camera sufficient or is it necessary to use a multispectral camera with five spectral bands? (v) What accuracy can be obtained when the previously trained empirical regression model is transferred to an independent site and environment? (vi) How many reference measurements are necessary for a user-based training of an empirical regression model at an independent site?

## **4.2 Material and methods**

### **4.2.1 Study site and experimental design**

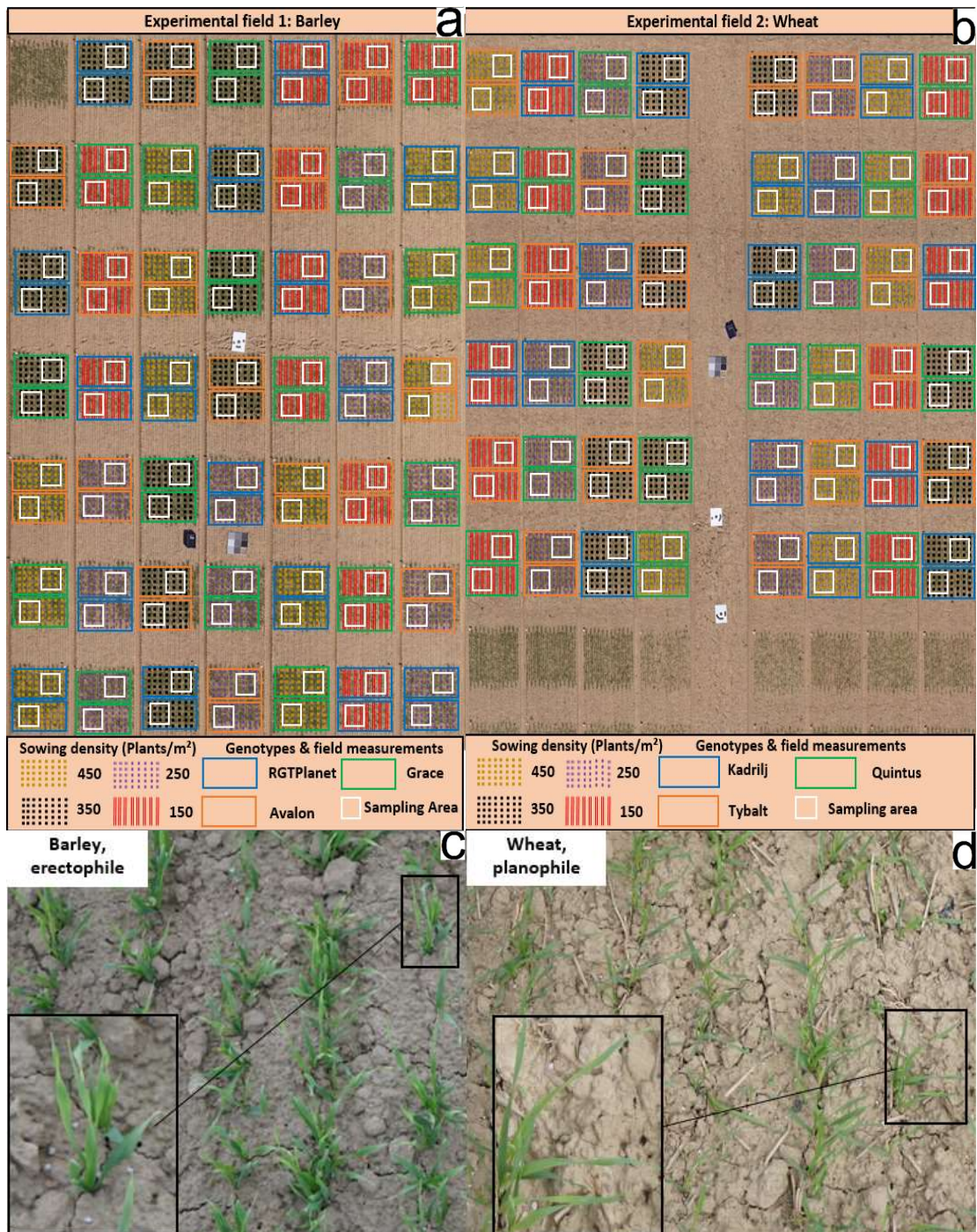
Three experimental fields were investigated on the research campus at the field lab on the Campus Klein-Altendorf. Two of them consisted of breeding plots ( $1.4 \times 3$  m) arranged in a randomized block design planted with monocultures of three spring barley genotypes (Avalon, Grace, RGTPlanet, Field 1, Figure 17a) and three spring wheat genotypes (Quintus, Kadrij, Tybalt, Field 2, Figure 17b).

The genotypes were sown in four different densities (150, 250, 350 and  $450 \text{ seeds m}^{-2}$ ) with eight repetitions (Figure 17a,b), resulting in 96 samples per species. Each repetition was divided into a sampling unit (SU) of one square meter. The seeds were sown in rows with a space of 10.4 cm for barley and wheat on 09 April. The experiments were treated with 80 kg N/ha on 19 April, herbicides were applied after data acquisition.

Spring wheat and spring barley differed strongly in their leaf arrangement. The leaf orientation of barley was erectophile with mainly vertical leaves, whereas wheat in this study had a planophile orientation with mainly horizontal leaves (Figure 17c,d).

Experimental field 3 was a breeding trial with 42 winter wheat genotypes arranged in a randomized block design ( $1.4 \times 3$  m). Three repetitions of each genotype were sown with a density of  $460 \text{ seeds m}^{-2}$  in rows with a space of 11 cm on 16 October ( $n=126$ ). Fertilization as well as herbicides were applied after data acquisition and do not have an influence on the scene in this study. This third experimental field was used to test the repeatability and practicability of the methodology. The experiment differed in genotype, sowing density, row spacing, sowing date, and season. Furthermore, the moisture content and thus the color of the soil varied among the experimental fields. The soils of experimental field 1 and 2 were dry and thus characterized by a light brown color. During wintertime, the soil of experimental site 3 was much darker because of the higher soil moisture.

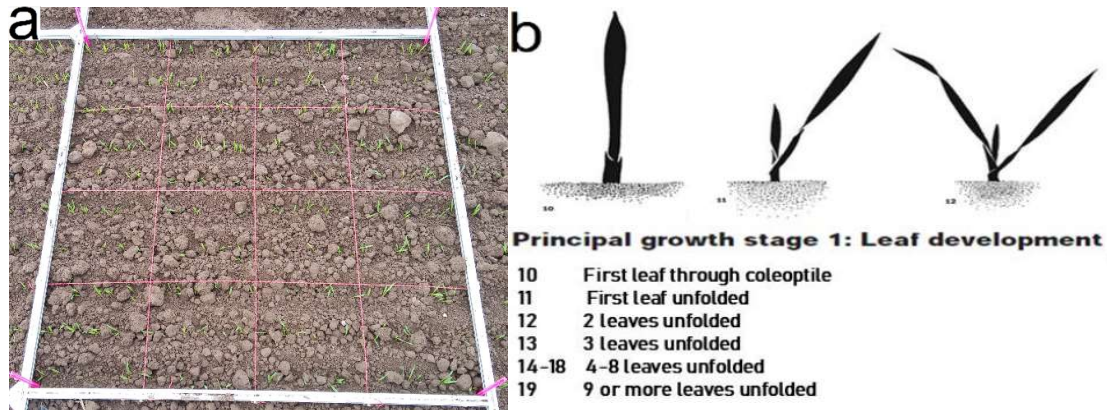




**Figure 17** Experimental fields 1 and 2, consisting of breeding plots with four sowing densities, three genotypes, and a different leaf arrangement of each species. **(a)** Experimental field 1: spring barley; **(b)** Experimental field 2: spring wheat; **(c)** Barley plants with mainly erectophile leaf arrangement, and **(d)** Wheat plants with mainly planophile leaf arrangement.

## 4.2.2 Field validation

For all plots in the three experimental fields, the individual plant numbers were counted in the field within one square meter. A metal frame covering one square meter was used to facilitate high plant counting accuracy in the field (Figure 18a). The corners were marked with pink sticks to highlight the area of interest for the UAV image data analysis.



**Figure 18 (a)** Example of the field validation setup. Metal frame (1x1 m) used for field plant counting with pink sticks at the corners to facilitate the identification of the region of interest in the UAV data; **(b)** Graphical illustrations of the principal growth stages in the leaf development of cereal plants according to the scale of the Biologische Bundesanstalt, Bundessortenamt und Chemische Industrie (BBCH) (modified according to Meyer (2001)).

Plant counting in the field was conducted according to the Biologische Bundesanstalt, Bundessortenamt und Chemische Industrie (BBCH) scale (Meier 2001) at stage 11 when the first leaf was unfolded and the tip of the second leaf became visible (Figure 18b). BBCH 11 has two main advantages as a reference point. First, the developmental stage enables to consider delayed development of seeds (BBCH 10) during the reference measurement. Second, at this stage the plant only consists of one unfolded leaf and therefore it is still possible to separate the individual plants, even if the seeds have germinated closely side by side.

## 4.2.3 UAV data acquisition

In this study, data were acquired with the MicaSense RedEdge multispectral camera. Image data at an flight altitude (FA) of 10 m above ground level resulted in a spatial resolution of 0.69 cm. In addition, the Sony Alpha RGB camera was used with a 20 mm fixed focal length lens, resulting in a spatial resolution of 0.20 cm at a FA of 10 m above ground level.

The experimental fields 1 and 2 were covered from these sensors four times. Table 8 provides an overview of the different acquisition dates, growth stages, flight altitudes, and illumination conditions. At the beginning of data acquisition, barley was already at BBCH stage 12, while wheat was at stage 11. For the subsequent acquisition dates, the plants were at the next BBCH stage (Table 8). For both crops, we recorded data at BBCH stages 12, 13,

and 14, which were expected to be the most appropriate growth stages for plant density assessment.

**Table 8** Data acquisition of the experimental fields 1 (barley) and 2 (wheat) with information on BBCH stages of the plants, sensors used, flight altitude, and illumination conditions for the data point. BBCH: Biologische Bundesanstalt, Bundessortenamt und Chemische Industrie; S: Direct solar illumination during data acquisition; (S): Minor clouds but predominantly direct solar illumination; O: Overcast sky; (O): Short periods of direct solar illumination but mainly overcast sky.

Nr.	Date	BBCH Stage		Sensor	Flight altitude (m)	Illumination conditions	
		Barley	Wheat			Barley	Wheat
1	22.04.2018	12	11	RGB	10	(S)	
					30	O	S
				Multispectral	10	(S)	
					20	S	
2	26.04.2018	13	12	RGB	10	O	(O)
					30		O
				Multispectral	10	(O)	S
					20	O	O
3	30.04.2018	14	13	RGB	10	S	
					30		O
				Multispectral	10	O	
					20		
4	04.05.2018	15	14	RGB	10		
					30		S
				Multispectral	10		
					20		

In addition to the different spectral and spatial properties of the RGB and multispectral images, data were acquired from two different altitudes with both cameras. This allowed us to investigate which spatial and spectral resolutions were optimal for each BBCH stage. The FAs of both cameras and the corresponding GSD are summarized in Table 9.

**Table 9** Spatial resolution of the UAV images depending on the sensor and the flight altitude.

Sensor	RGB		Multispectral	
	Flight altitude (m)	10	30	10
Ground sampling distance (cm/pixel)	0.20	0.59	0.69	1.38



Furthermore, illumination conditions varied between direct solar illumination, completely overcast sky and the combination of both conditions at the different dates of data acquisition (Table 8). The first and fourth data sets were acquired under predominantly direct solar, while the second and third data sets were recorded under predominantly overcast conditions.

For experimental field 3, a data set was acquired on 17 November, at BBCH stage 12, with the multispectral sensor from an FA of 20 m above ground level (GSD = 1.38 cm) under overcast conditions. For these data acquisition parameters, flight altitude and sensor were selected based on the results of the best-performing wheat empirical regression model in relation to the lower GSD.

#### 4.2.4 Data analysing

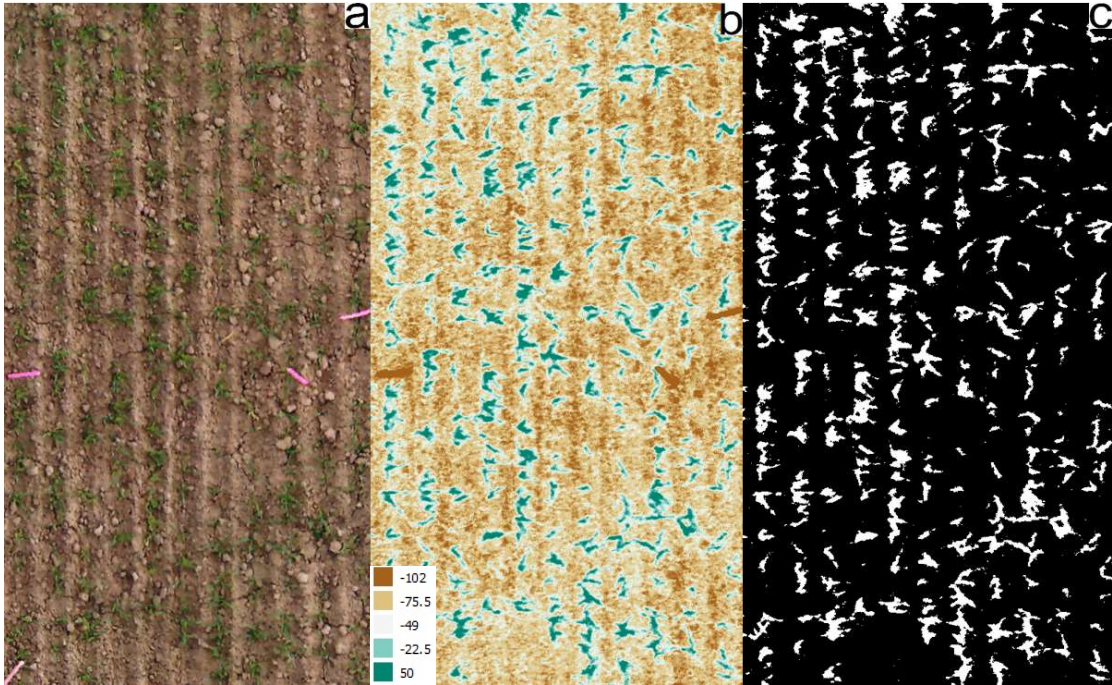
For the analysis in this study, the UAV images were processed to orthomosaics using Agisoft Metashape software (version 1.5.5) take into account the necessary processing settings (cf. Table 3).

The excess green minus excess red (ExGR) (Neto 2010) vegetation index was calculated for the RGB camera data (Equation 3) and the normalized difference vegetation index (NDVI) (Tucker 1979) for the multispectral data (Equation 4).

$$ExGR = (2 \textit{Green} - \textit{Red} - \textit{Blue}) - (1.4 \textit{Red} - \textit{Green}) \quad (3)$$

$$NDVI = \frac{Ref_{NIR} - Ref_{RED}}{Ref_{NIR} + Ref_{RED}} \quad (4)$$

The vegetation indices finally allowed the fractional cover assessment by applying a threshold that divided the pixels into two classes; foreground (plant pixels) and background (soil pixels). The threshold was automatically determined by the variance between two classes based on the Otsu method (Otsu 1979). The use of Otsu thresholding is an established method based on the aforementioned variance between two classes to distinguish vegetation from the background (Torres-Sánchez et al. 2015; López-Granados et al. 2016; Bassine et al. 2019; B. Li et al. 2019; Marcial-Pablo et al. 2019). The procedure and segmentation performance are exemplarily illustrated in Figure 19. Shapefiles based on the aforementioned SU were finally used to calculate the fractional cover (number of plant pixels per square meter) for the calculated vegetation indices (ExGR, NDVI) in the processed orthomosaics.



**Figure 19** (a) Section of an UAV RGB orthomosaic at BBCH stage 13; (b) Calculated ExGR in the region of interest, and (c) Segmentation of plants (white) and background (black) using the ExGR and Otsu thresholding.

#### 4.2.5 Statistical analysis

The manually counted plant density in combination with the determined fractional cover per SU allowed to develop linear empirical regression models for plant density assessment. For this, a k-fold cross-validation was used to estimate the model and test the error rate.

Firstly, we randomly divided the data sets of experimental fields 1 and 2, each with a sample size of 96, 48 times into a calibration and validation data set for the empirical regression model development. Two-thirds were used to train the model (calibration), whereas the remaining third was used to evaluate the model performance (validation). Across the models, we determined the performance using R (Core team 2018) based on the averaged coefficient of determination ( $R^2$ ), root mean square error (RMSE) and mean absolute error (MAE). The advantage of the MAE over the RMSE is that it gives the same weight to all errors. Therefore, the MAE is more robust to outliers. In contrast, the commonly used RMSE weights errors with larger absolute values more than errors with smaller absolute values and thus is more sensitive to outliers (Willmott and Matsuura 2005; Chai and Draxler 2014).

In order to identify which settings provide the highest model accuracies and smallest errors, three-way ANOVAs were conducted to verify the MAE including a Tukey post-hoc test. ANOVAs were fitted for each species separately. The following model in R notation was used:  $MAE \sim (\text{Vegetation Index} * \text{BBCH} * \text{GSD})$ .

Secondly, a previously calibrated empirical regression model obtained from experimental field 2 was transferred to the third experimental field according to Equation 5. In this context, the best-performing wheat model for the low GSD imagery and multispectral camera was chosen (Table 11, BBCH 12, NDVI, 20 m; Slope = 0.171, Intercept = -15.628). By transferring a previously calibrated empirical regression model to an independent data set with a different experimental design (genotypes, sowing densities) and environmental conditions (season, soil color) it was possible to verify the model robustness and the repeatability of the methodology at the same time.

$$UAV \text{ estimated plant density} = \frac{\text{fractional cover} - \text{intercept}}{\text{slope}} \quad (5)$$

In addition, we again randomly divided the data, consisting of 126 samples, 63 times into a calibration and validation data set for the third experimental field. Each time either 1/2, 1/3, 1/4, 1/6, 1/8, 1/11, 1/15 or 1/25 of the total number of samples was used for model calibration. Thus, it was possible to identify the minimum number of reference measurements required to build up a user-based empirical regression model adapted to a specific experimental layout and environment. Each time, the remaining part of the data set was used to validate the performance of the model based on the averaged R<sup>2</sup>, RMSE, and MAE. To identify significant deviations between the different sample sizes used for the model calibration, one-way ANOVAs, including a Tukey post-hoc test, were conducted to verify the MAE.

## 4.3 Results

### 4.3.1 Experimental field 1: Barley

The results of all individual models predicting the plant density of spring barley from UAV fractional cover values are summarized in Table 10. It can be seen that the spatial resolution had a substantial impact on ExGR model performance at the early stage of leaf development with only two unfolded leaves per plant (BBCH 12). The high GSD (0.20 cm) provided a model performance with an  $R^2$  of 0.92 and an MAE of 26 plants  $m^{-2}$ . With decreasing GSD (0.59 cm), the correlation declined ( $R^2 = 0.70$ ) with regression error metrics (RMSE, MAE) of more than 55 plants  $m^{-2}$  (Table 10).

**Table 10** Statistical relationship between manually counted plant density and the fractional cover of barley at different BBCH stages, VIs, FAs, and GSDs. Averaged statistical measures were computed across the empirical regression models randomly divided into calibration and validation data sets. Different letters indicate significant differences between MAE ( $p < 0.05$ ). VI: vegetation index; FA: flight altitude; GSD: Ground sampling distance; Ill: Illumination conditions;  $R^2$ : averaged coefficient of determination; RMSE: averaged root mean square error; MAE: mean absolute error; ExGR: excess green red; NDVI: normalized difference vegetation index.

BBCH	VI	FA (m)	GSD (cm)	Ill.	Calibration (n = 64)			Validation (n = 32)				
					Slope	Intercept	$R^2$	Slope	Intercept	$R^2$	RMSE	MAE
12	ExGR	10	0.20	(S)	0.051	-1.666	0.91	1.010	-0.553	0.92	34	26ab
		30	0.59	O	0.085	-6.632	0.68	1.031	-6.796	0.70	75	55h
	NDVI	10	0.69	(S)	0.089	-10.107	0.93	1.017	-5.217	0.92	32	24a
		20	1.38	S	0.105	-15.201	0.89	1.009	-3.436	0.90	37	29bc
13	ExGR	10	0.20	O	0.077	-1.556	0.92	1.007	-3.432	0.92	34	26ab
		30	0.59	O	0.139	-12.822	0.87	1.001	-0.766	0.89	40	31c
	NDVI	10	0.69	(O)	0.108	-7.756	0.90	1.010	-1.392	0.90	38	26ab
		20	1.38	O	0.153	-15.301	0.90	1.012	-1.727	0.90	37	26ab
14	ExGR	10	0.20	S	0.113	0.422	0.80	0.984	4.957	0.78	56	44g
		30	0.59	S	0.177	-22.158	0.84	0.987	5.034	0.83	48	34d
	NDVI	10	0.69	O	0.130	-3.319	0.85	1.018	-4.930	0.85	47	35d
		20	1.38	O	0.162	-8.280	0.83	1.022	-5.527	0.84	49	38ef
15	ExGR	10	0.20	S	0.111	24.883	0.84	0.998	0.789	0.83	50	41f
		30	0.59	S	0.147	21.045	0.82	0.999	-2.178	0.83	50	39ef
	NDVI	10	0.69	S	0.163	4.893	0.85	0.996	0.970	0.86	45	36de
		20	1.38	S	0.184	5.947	0.84	1.001	-0.997	0.85	47	39ef

In contrast, the spatial resolution had less impact on model performance at BBCH stage 12 using the NDVI to predict the plant density. The results of the low GSD (1.38 cm) showed only slightly lower model performance ( $R^2 = 0.90$ , MAE = 29 plants  $m^{-2}$ ) in comparison to

the model constructed with the high GSD data (0.69 cm). This model provided the best performance with the highest  $R^2$  of 0.92 and an MAE of 24 plants  $m^{-2}$  (Table 10).

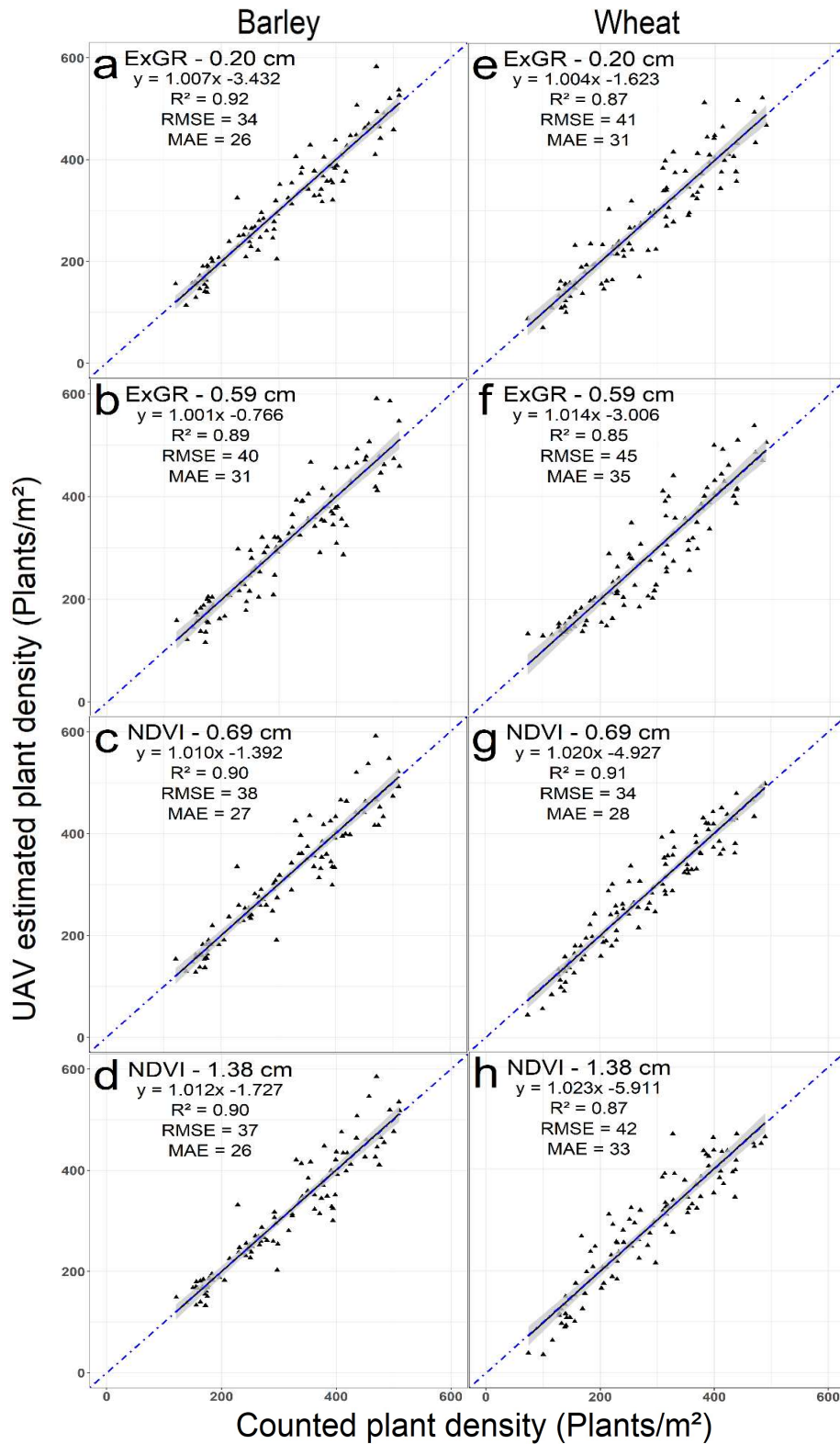
At BBCH stage 13, the best validation performance was observed for the ExGR with uncertainties of less than 10 % for both GSDs (GSD = 0.20 cm,  $R^2 = 0.92$ , MAE = 26; GSD = 0.59 cm,  $R^2 = 0.89$ , MAE = 31), also illustrated in Figure 20. Moreover, BBCH stage 13 was as well the optimal development stage for the multispectral data and low GSD (NDVI, 0.69 cm), which led to an  $R^2$  of 0.90 and MAE of 26 plants  $m^{-2}$  (Table 10). The regression line showed a good match to the 1:1 line, with almost no over- or underestimation of the predicted plant density (Figure 20, a-d).

The model performance of both vegetation indices declined in the subsequent growth stages (BBCH 14, 15), where the leaves grow in size and the plants develop further leaves. The  $R^2$  of the regression models, varying between 0.78 and 0.86, was still high, but the regression error metrics (RMSE, MAE) of 34-44 plants  $m^{-2}$  indicated lower model accuracy (Table 10). In general, a higher BBCH stage led to a higher slope in the calibration models (Table 10). This was influenced by the fact that areas with a higher number of plants have a proportionally stronger increase in fractional cover over time compared to areas with a lower number of plants. The slopes in the validation models with values around one and intercepts of around zero for almost all BBCH stages demonstrated the high prediction accuracy of the regression models.

### 4.3.2 Experimental field 2: Wheat

The model performance at early leaf development stages of spring wheat (BBCH 11, 12) was also influenced by the spatial resolution. The ExGR calculated from the high GSD (0.20 cm) data led to a significantly higher prediction accuracy compared to the models based on the low GSD (0.59 cm) data (Table 11). The spatial resolution has less impact on model performance when the NDVI was used for plant density assessment. The NDVI provided similar model accuracies for data sets recorded from different GSDs within a BBCH stage. Even at BBCH stage 12, the model for multispectral data and low GSD imagery (NDVI, 1.38 cm) already led to the best performance ( $R^2 = 0.89$ , MAE < 29 plants  $m^{-2}$ ) with uncertainties of less than 10 % (Table 11). However, it can be observed that the model made use of the image data with a lower GSD (1.38 cm) led to better results compared to the model based on data with the higher GSD (0.69 cm). Considering the illumination conditions, the high GSD data was acquired under sunny, while the low GSD data was recorded under cloudy conditions.

Almost the same prediction accuracy was observed for the NDVI model constructed using the low GSD data (1.38 cm) at BBCH stage 13 (Figure 20), while the higher GSD model provided the highest  $R^2$  of 0.91 and the lowest MAE of 28 plants  $m^{-2}$  of all models (Table 11). The best model performances for the ExGR and both GSDs were also observed when the third leaf of the plants was unfolded (BBCH 13). Both models had an  $R^2$  higher than 0.85 and MAE of less than 35 plants  $m^{-2}$  (Figure 20).



**Figure 20** Linear regression between manually counted and UAV estimated plant density for different vegetation indices and GSDs obtained for BBCH stage 13 (n = 96). The UAV estimated plant density represents an averaged number of plants computed by the randomly divided models for (a-d) barley and (e-h) wheat. Black lines represent regression lines with 95 % confidence intervals; blue lines represent 1:1 line.



Similar to barley, the regression lines showed a good match with the 1:1 lines and only less scattering was visible (Figure 20, e-h). At BBCH stage 14, under sunny conditions, the accuracy for the ExGR models clearly decreased. In contrast, the NDVI models led to similar performances as at the BBCH stage 13 (Table 11). As already observed for barley, higher BBCH stages led to a higher slope value for the calibration models (Table 11). In contrast, these slope values were higher for wheat compared to barley (Table 10). The slopes of the validation models had values close to one, which illustrated the high prediction accuracy of the generated models (Table 11). However, the performance evaluation for wheat led to a slightly lower R<sup>2</sup> and higher regression error metrics compared to barley.

**Table 11** Statistical relationship between manually counted plant density and the fractional cover of wheat at different BBCH stages, VIs, and FAs. Averaged statistical measures were computed across the empirical regression models randomly divided into calibration and validation data sets. Different letters indicate significant differences between MAE ( $p < 0.05$ ).

BBCH	VI	FA (m)	GSD (cm)	Ill.	Calibration (n = 26)			Validation (n = 32)				
					Slope	Intercept	R <sup>2</sup>	Slope	Intercept	R <sup>2</sup>	RMSE	MAE
11	ExGR	10	0.20	S	0.029	1.394	0.81	1.007	-3.532	0.81	52	40fg
		30	0.59	S	0.036	-1.706	0.59	1.007	-4.064	0.60	87	67j
	NDVI	10	0.69	S	0.086	-7.358	0.83	1.000	0.414	0.83	50	39fg
		20	1.38	S	0.078	-11.406	0.80	1.010	-2.157	0.81	54	42g
12	ExGR	10	0.20	(O)	0.082	-8.013	0.83	1.013	-4.587	0.84	48	38ef
		30	0.59	O	0.089	-14.445	0.63	1.039	-11.496	0.63	86	62i
	NDVI	10	0.69	S	0.116	-5.648	0.84	1.011	-4.485	0.85	45	34cd
		20	1.38	O	0.171	-15.628	0.89	1.004	-3.493	0.89	37	29ab
13	ExGR	10	0.20	O	0.148	-4.750	0.87	1.004	-1.623	0.87	41	31bc
		30	0.59	O	0.219	-26.645	0.86	1.014	-3.006	0.85	45	35de
	NDVI	10	0.69	O	0.160	4.893	0.91	1.020	-4.927	0.91	34	28a
		20	1.38	O	0.182	6.681	0.87	1.023	-5.911	0.87	42	33bc
14	ExGR	10	0.20	S	0.103	17.981	0.83	0.960	10.483	0.82	50	38fg
		30	0.59	S	0.133	16.955	0.78	0.956	10.650	0.76	59	46h
	NDVI	10	0.69	S	0.180	1.230	0.90	1.019	-4.790	0.90	36	28a
		20	1.38	S	0.200	9.290	0.86	1.017	-4.812	0.87	43	34cd

### 4.3.3 Ideal parameters for data acquisition

The three-way ANOVA for the MAE was calculated for barley (Table S1, Table S2) and wheat (Table S3, Table S4) to identify the ideal parameters for plant density assessment and data acquisition. In the case of barley the vegetation index, the BBCH stage, and the interaction between BBCH and GSD explained most of the variation in the methodology (Table S1). The ANOVA results identified in general the NDVI as the vegetation index,

BBCH 13 as the development stage and the high GSD as the ideal parameter for data acquisition and plant density assessment of barley (Table S2)

For wheat, the vegetation index clearly explained most of the variation (Table S3). The GSD also had a high impact on the plant density assessment. The interaction between BBCH and GSD was the third most important factor. Despite the investigated differences of barley and wheat, the ANOVA of wheat identified the same predictors (NDVI, BBCH 13, high GSD) for data acquisition and plant density assessment (Table S4).

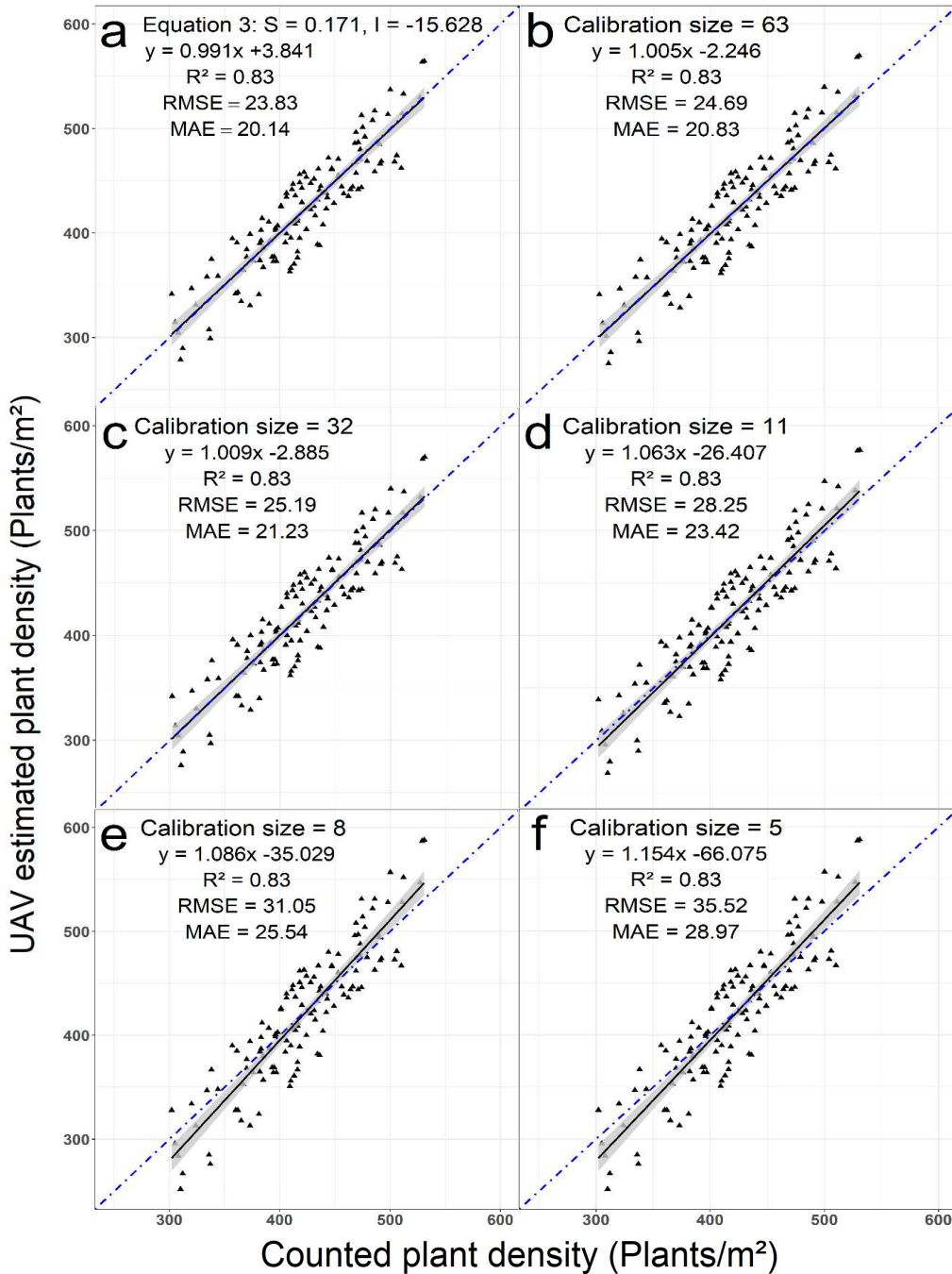
#### **4.3.4 Experimental field 3: Investigate repeatability**

In the case of the third experiment, two investigations were conducted to evaluate the practicability and repeatability of the developed methodology. Firstly, the best-performing empirical regression model for wheat and the low GSD (Table 11, BBCH 12, NDVI, 20 m; slope = 0.171, intercept = -15.628) was applied to the data of experimental field 3 according to Equation 5. The best-performing model of the lower GSD was considering instead of the ideal model performance to increase the practicability of the methodology. The experiments differed with regard to season, soil color, genotypes, and sowing density. Especially with its 42 genotypes and the large number of variations in plant characteristics, the third experimental field was well suited for testing the repeatability and practicability of the previously developed regression model.

The transfer of the calibrated empirical regression model for UAV based plant density prediction yielded in a high prediction accuracy. The determined MAE was lower than 20 plants  $m^{-2}$  and the  $R^2$  of 0.83 was relatively high (Figure 21a). The regression line showed a good match with the 1:1 line and only lower values were slightly overestimated.

Secondly, the size of the calibration data subset was reduced stepwise to determine the minimum number of field measurements necessary for a user-based calibration of a new empirical regression model (Table 12). The starting calibration subset consisting of 63 samples also provided the highest accuracy with an MAE of less than 21 plants  $m^{-2}$  and an  $R^2$  of 0.83. The prediction accuracy of the user-based calibration model based on the 63 training samples and the corresponding scatterplot (Figure 21b) were almost similar compared to the transferred calibrated empirical regression model from experimental field 2 (Figure 21a).

The continuous reduction of the calibration sample size from 63 to 11 samples led to no significant changes in model performance (Table 12). The slope and intercept values of the validation regression functions slightly increased with decreasing size of the calibration samples. However, the residual deviations and the 95 confidence intervals were quite uniformly independent of the calibration sample sizes (Figure 21b-d). The last two reduction steps from 11 to only eight and five samples led to significantly lower model performances (MAE > 26 plants  $m^{-2}$ ) (Table 12). Additionally, the corresponding scatterplots (Figure 21e,f) illustrated an underestimation of lower values and an overestimation of higher values.



**Figure 21** Linear regression between manually counted wheat plant densities and those estimated from UAV data at BBCH stage 12 for experimental field 3 with two investigations (n = 126): **(a)** The empirical regression model from experimental field 2 was transferred to the data of experimental field 3 and **(b-f)** Investigation of the necessary number of field measurements for a user-based calibration of an empirical regression model based on stepwise reduction of calibration size to **(b)** 63 samples; **(c)** 32 samples; **(d)** 11 samples; **(e)** 8 samples, and **(f)** 5 samples. The UAV estimated plant density represents an averaged number of plants computed by the randomly divided models. Black lines represent regression lines with 95 % confidence intervals; blue lines represent 1:1 lines. S: Slope; I: Intercept.

**Table 12** Analysis results of the number of field measurements necessary to calibrate an empirical regression model to a specific layout or environment (n = 126). The calibration sample size was reduced stepwise from 63 to five reference measurements, while the remaining samples were used to validate the performance of the calibrated models. Averaged statistical measures were computed across the empirical regression models randomly divided into calibration and validation data sets for each sample size. Different letters indicate significant differences between MAE ( $p < 0.05$ ).

Experimental field 3									
Calibration				Validation					
n	Slope	Intercept	R <sup>2</sup>	n	Slope	Intercept	R <sup>2</sup>	RMSE	MAE
63	0.169	-14.868	0.83	63	1.005	-2.246	0.83	25	21a
42	0.169	-14.955	0.83	84	1.008	-2.976	0.83	25	21a
32	0.169	-14.972	0.82	94	1.009	-2.885	0.83	25	21a
21	0.168	-14.446	0.81	105	1.024	-10.355	0.83	26	22a
16	0.168	-14.441	0.81	110	1.026	-11.117	0.83	27	22a
11	0.165	-12.837	0.81	115	1.063	-26.407	0.83	28	23ab
8	0.165	-13.404	0.78	118	1.086	-35.029	0.83	31	26bc
5	0.161	-11.325	0.79	121	1.154	-66.075	0.83	36	29c

## 4.4. Discussion

In the study presented here, UAV data were evaluated for high-throughput field phenotyping of plant density. The results demonstrated that plant density could be predicted with potential uncertainties of less than 10 %. In the following, the introductory questions are discussed in detail.

BBCH stage 13 was identified as the plant developmental stage best suited for UAV data acquisition (Table S1-S4). At this growth stage, the MAEs of the validation models were between 25 and 34 plants  $m^{-2}$  and therefore similar compared to previous studies focusing on wheat (Jin et al. 2017b; Liu et al. 2017a; Liu et al. 2017b). In addition, we observed a high prediction accuracy at BBCH stage 12 when plant density prediction was based on the data recorded from the higher GSD ( $\emptyset$  MAE < 30 plants  $m^{-2}$ ) (Table 10, 11). In summary, the plants or the resolution needs to be large enough to detect the fractional cover accurately. Even with four unfolded leaves per plant (BBCH 14) predictions of plant density were possible with MAEs ranging from 28 to 46 plants  $m^{-2}$ . However, the model accuracy partly declined at BBCH stage 14 and 15 compared to the previous stages. The leaves grow in size and the plants develop additional leaves (Figure 18b), resulting in a higher overlap between neighboring plants.

In other studies, the investigated methodology was limited to a specific growth stage (Sankaran et al. 2015; Liu et al. 2017a; Jin et al. 2017b; Liu et al. 2017b; Aich et al. 2018). The approach presented here, however, allowed plant density to be estimated from UAV data acquired during BBCH stages 12 to 14. This makes the proposed method more practical, since UAV data acquisition is possible in a longer time window.

The models constructed using the higher GSD data enabled a higher prediction accuracy of plant density. However, especially at the later stages of plant development (BBCH 13 onwards), the differences in the predictions between the models for the GSDs were not great. At BBCH stage 13, the averaged MAE was lower than 31 plants  $m^{-2}$  for the low GSD image data (ExGR-30 m, NDVI-20 m) and had an almost similar prediction accuracy compared to the high GSD (ExGR-10 m, NDVI-10 m,  $\emptyset$  MAE < 28 plants  $m^{-2}$ ). In principle, the prediction accuracy was influenced less by the spatial resolution using the NDVI for plant density assessment (Table 10, Table 11). We conclude that multispectral data recorded with a lower GSD than 1.38 cm may predict a similar accuracy to increase the throughput of the approach.

In comparison to other studies (Liu et al. 2017a; Jin et al. 2017b; Liu et al. 2017b; Aich et al. 2018), the present methodology is not limited to spatially very highly resolved image data (0.02 cm) in order to obtain precise estimates of plant density. The results showed that for ideal conditions RGB data with 0.6 cm and multispectral data with 1.4 cm spatial resolution were sufficient to predict the plant density of numerous plots, greatly enhance the throughput of the approach. In addition, the automatic image acquisition of the UAV and the use of orthomosaics instead of individual images (Liu et al. 2017a; Liu et al. 2017b; Aich et al.

2018) are ideally for large scales and decision-making process in precision farming and breeding.

Cereals such as wheat and barley have comparable plant structures, especially at the early leaf developmental stages. However, some individual characteristics such as the leaf arrangement can be different. The results stressed that it was essential to distinguish between planophile (wheat) and erectophile (barley) leaves in the process of developing empirical regression models having the capability to estimate plant density precisely. In a nadir UAV image, a leaf with horizontal orientation (planophile) covers a larger soil area and consists of a higher number of green pixels in comparison to a leaf with vertical orientation (erectophile) (Figure 17 c,d). This results in higher fractional cover in the case of the observation of planophile cereals. For that reason, the wheat calibration models had steeper slopes compared to the barley calibration models. The impact of the leaf arrangement on the methodology was not investigated in previous studies focusing on the assessment of plant density (Sankaran et al. 2015; Liu et al. 2017a; Jin et al. 2017b; Liu et al. 2017b; Aich et al. 2018; Koh et al. 2019).

The comparison of the sensors revealed that the multispectral camera had a higher prediction accuracy compared to the RGB sensor (Table S2, Table S4). Especially for the prediction of wheat plant density, the vegetation index was the most important factor (Table S3). The predominantly horizontally oriented leaves of wheat caused more mixed pixels. The additional near- infrared spectral band used for the calculation of the NDVI is advantageous for distinguishing plant from soil pixels. Furthermore, the radiometric correction of the multispectral data based on the solar irradiance sensor and reference panel enabled to convert the raw pixel information into absolute spectral reflectance. This made the method more robust to varying illumination conditions among data sets and during data acquisition (Table 8). Based on these findings, multispectral data are recommended in particular for genotypes with planophile leaves.

In order to test the repeatability and robustness of the approach, in particular with regard to genotypic varieties, the calibrated empirical regression model constructed for spring wheat was transferred to an independent breeding experiment with 42 different winter wheat genotypes. Visual field observation revealed only little variability among genotypes in the early leaf developmental stages. This observation was similar to the findings of Jin et al. (2017b). The transferred regression model, therefore, provided good prediction accuracy across the genotypes ( $R^2 = 0.83$ ,  $MAE < 20 \text{ plants m}^{-2}$ ). The large variation in the training dataset resulted in a robust model, which avoids the need for further genotypic model adaptation. For this reason, a specific re-calibration, which is usually required for machine-learning techniques (Jin et al. 2017b; Liu et al. 2017b; Koh et al. 2019), was not obligatory using this approach and comparable conditions in this study.

Estimating the required number of reference measurements in the field to calibrate a user-based empirical regression model is needed in the case of different seed row spacing, illumination conditions and different spatial resolution of the UAV data. For this user-based calibration, 11 randomly selected field measurements were already sufficient in this study to



train a robust model ( $MAE < 23 \text{ plants m}^{-2}$ ). No significant differences in prediction accuracy were observed by using a small training subset of 11 samples instead of a larger one consisting of 63 samples (Table 12, Figure 21). Nevertheless, attention has to be paid to outliers when such a small calibration sample size is used. Moreover, some variation of plant density in the field, which covers a range of possible values, is needed to adequately train a user-based empirical regression model. The required time to manually count the plants in an area of one square meter with a sowing density of  $460 \text{ seeds m}^{-2}$  is approximately 5 minutes.

In summary, the prediction accuracy of the transferred regression model (Figure 21a) was closely related to the user-based calibrated model (Figure 21b-e), although the experiments differed by season, soil color, genotypes, and sowing density. Therefore, we assumed that user-based calibration is not necessary as long as the conditions are comparable, such as seed row spacing, the spatial resolution of the UAV data and the illumination conditions during data acquisition. In this respect, it is important to note that the empirical regression model was calibrated and transferred with data acquired under overcast conditions. Hence, the transferability of the model was not affected by the position of the sun. It is important to note that a calibrated model should be transferred to data with comparable illumination conditions. This is an important prerequisite for the transferability of empirical regression models (Dorigo et al. 2007). Therefore, we recommend that data should be primarily acquired under overcast conditions or as an alternative timely close to solar noon to enable high transferability of the calibrated model. Under direct sunlight, data acquisition around solar noon conditions has the benefit of minimizing shadow effects and enabling comparable illumination conditions between datasets.

The investigation of repeatability focused on wheat with predominantly planophile leaves, higher regression error metrics and high sowing density of up to  $500 \text{ plants m}^{-2}$  to demonstrate the transferability and robustness of the calibrated regression model. It can be expected that the model performance increases for predominately erectophile leaves and lower sowing densities. However, the prediction capability in this study was not negatively affected at early leaf developmental stages by high sowing densities of up to 500 plants with a higher probability of overlapping leaves (Jin et al. 2017b; Liu et al. 2017a; Liu et al. 2017b). However, a weed-free crop stand is required to achieve a sufficient prediction accuracy, since weeds have a negative impact on the assessment of fractional cover. Furthermore, local soil heterogeneities or sowing failures need to be ideally identified and subsequently considered to enable a successful differentiation among varieties.

## 4.5 Conclusions

The study presents a novel way to assess the plant density of cereal crops with high-throughput UAV image data. The determination of fractional cover for plant density assessment takes advantage of the fact that the average plant size is relatively stable and the number of overlapping plants is relatively low at early leaf developmental stages. BBCH stage 13 was identified as the ideal growth stage when the plants were large enough to detect the fractional cover accurately even with lower GSDs in this study. Nevertheless, with appropriate parameters such as a high GSD for BBCH stage 12, the specific growth stage was not critical for UAV based plant density assessment.

The calibrated models proved to be robust with respect to unknown sites with comparable conditions such as seed row spacing, spatial resolution, and illumination conditions. Across a data set of 42 genotypes, it was shown that genotypic model adaptation is not necessary. It is only essential to distinguish between planophile and erectophile cereals in the application and process of empirical regression models. Better results were observed when multispectral data were used for plant density quantification, in particular for planophile cereals. However, also RGB data provided comparable model performance and was sufficient for plant density assessment.

The results stressed the transferability, practicability, and repeatability of the developed methodology. With conditions comparable to the presented study, the operator can choose whether to apply the trained regression model without re-calibration or whether to perform a user-based approach with a small number of field measurements. Furthermore, also a broader application of the method is conceivable, e.g. for other cereals with heterogeneous germination and overlapping leaves such as oats or triticale. In contrast to machine-learning approaches, expert guidance in feature extraction and large training datasets are not necessary.

The developed methodology facilitates high-throughput plant density assessment of cereal crops for decision-making in precision farming and breeding. Breeding trials are particularly difficult to monitor within a reasonable time, resulting in an increasing need for faster selection of superior lines. In addition, considering the plant density variation caused, for example, by different soil or nutrition conditions, can increase farmer's crop production and yield estimations.

## 5. Study III: Spike density quantification of different wheat varieties from UAV RGB imagery

Norman Wilke <sup>1,\*</sup>, Bastian Siegmann <sup>1</sup>, Mark Müller-Linow<sup>1</sup>, Erekle Chakhvashvili <sup>1</sup> and Uwe Rascher <sup>1</sup>

<sup>1</sup> Institute of Bio- and Geosciences, Plant Sciences (IBG-2), Forschungszentrum Jülich GmbH, 52428 Jülich, Germany

\* Correspondence: n.wilke@fz-juelich.de; Tel.: +49 2461 61-1733

Status: In preparation for submission

**Author Contributions:** Norman Wilke: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. Bastian Siegmann: Conceptualization, Supervision, Visualization, Writing – original draft, review & editing. Mark Müller-Linow: Software, Supervision, Writing – review & editing. Erekle Chakhvashvili: Writing – review & editing, Uwe Rascher: Funding acquisition, Project administration, Supervision, Resources.

### Abstract

The real-time information regarding agronomic traits, such as spike density can enhance the quality of yield estimations, variety selection, and agricultural management. At present, spikes are still quantified by subjective and expensive manual labor, without a standard protocol for counting. Different image processing techniques, algorithms, and machine learning approaches have already been applied to increase the throughput and objectivity of spike quantification in-field conditions. However, the methods relied mainly on ground-based systems and spatially, highly-resolved images. In addition, the adaptation to external conditions (e.g., illumination, sensors) or to specific phenotypic variations among genotypes remains a significant challenge. In this study, a method is presented that makes use of the statistical relationship between manually counted spike density and spike cover (number of spike pixels per square meter) derived from an RGB camera equipped on UAV. Three different wheat genotypes, with morphological variations, such as rate of awn occurrence, color and spike inclination were examined. With a GSD of 1.4 mm, the spike density could be accurately modelled ( $rRMSE < 15\%$ ,  $RMSE < 18 \text{ spikes m}^{-2}$ ) among the phenotypes in this study. Furthermore, the results of the power analysis led to the assumption that 11 reference measurements were appropriate for a user-based calibration of an empirical regression model. This would potentially allow for an adaption of the approach to genotypic specificities and the external conditions of the scene, which require only a small number of field measurements for high-throughput phenotyping of spike density.

## 5.1 Introduction

Wheat production plays a major role in ensuring global food security (Alexandratos and Bruinsma 2012). The correct selection of varieties for abiotic and biotic stresses has the potential to increase grain yield in the future (Slafer et al. 1996; Slafer et al. 2014). The number of spikes per unit of ground area (spike density) has been proven to correlate to wheat yield, together with a thousand corn weight and the number of grains. Therefore, spike density is a vital trait for early yield predictions (García Del Moral et al. 2003; Zhang et al. 2007; Pask et al. 2012; Ferrante et al. 2017). The current practice involves the manual counting of spikes within several small subsample areas to represent field or plot conditions. This requires an additional step of extrapolation to achieve comparable standard units ( $m^{-2}$ ), leading to unknown uncertainties, especially in the case of a high spatial variability in the distribution of spikes.

Several alternative approaches have been developed to quantify spikes in the field. Thermal image data enable to distinguish between spikes and canopy elements, due to a lower transpiration and the higher temperature of the spikes (Fernandez-Gallego et al. 2019). However, the slight temperature difference of the spikes in comparison to the rest of the canopy makes their detection challenging. Cloud cover, the angle of the sun, and changes in wind speed, possibly caused by UAV movements, are factors influencing the derived canopy surface temperatures (Jackson et al. 1983; Chapman et al. 2014) and can complicate spike detection.

Other approaches primarily make use of conventional, high-resolution RGB data. Information on texture, color, and morphology were utilized for spike quantification, e.g., using filtering techniques, grayscale transformation and algorithms (Cointault et al. 2008; Liu et al. 2014; Alharbi et al. 2018; Fernandez-Gallego et al. 2018; Zhou et al. 2018a). Subsample areas of highly-resolved images were selected, leading to promising results for spike identification. In addition, machine learning approaches like CNNs (Hasan et al. 2018; Madec et al. 2019; Sadeghi-Tehran et al. 2019) and SVM (Zhu et al. 2016; Zhou et al. 2018b) have been successfully trained to RGB data for spike recognition.

However, the data acquisition in relation to the aforementioned approaches was limited to ground-based systems like a camera platform (Cointault et al. 2008), self-built vehicles (Hasan et al. 2018; Zhou et al. 2018a), fixed-site phenotyping platforms (Zhu et al. 2016; Sadeghi-Tehran et al. 2019), a boom (Madec et al. 2019) or was acquired predominantly by hand (Liu et al. 2014; Alharbi et al. 2018; Fernandez-Gallego et al. 2018; Zhou et al. 2018b; Fernandez-Gallego et al. 2019; Sadeghi-Tehran et al. 2019). Moreover, for a sufficient spike detection, images were acquired from the side in some cases, at an angle of 45 degrees above the horizontal (Alharbi et al. 2018; Hasan et al. 2018; Zhou et al. 2018b), rather than from a nadir view.

These data collection requirements clearly limited throughput, automation capabilities, and applicability. Firstly, images from each sampling unit (SU) were manually acquired and analyzed individually, without any information regarding their geolocation on the field scale.

Secondly, the aforementioned selection of subsample areas required an additional step to extrapolate in standard units (spikes per square meter). Thirdly, methodologies were applied to highly-resolved images ( $GSD < 0.60$  mm) and small image footprints, to allow for good prediction accuracy. Madec et al. (2019) and Sadeghi-Tehran et al. (2019) have already demonstrated that a lower spatial resolution ( $GSD < 1.00$  mm) leads to a clearly decreased identification performance of spikes. The necessity for such a high resolution or image acquisition from the side distinctly limits the application possibilities of e.g., a UAV for data acquisition.

Recently, only Fernandez-Gallego et al. (2020) has successfully made use of UAV images for the spike density assessment, with machine learning techniques. However, the approach has been applied to very similar awnless genotypes and would be problematic for large scales. The UAV images were taken in manual mode to avoid different image settings and to minimize errors caused by distortion or perspective. Moreover, a manual selection of the plot centroids was necessary.

In general, developed methodologies are adapted and trained to external conditions like illumination, background, soil moisture content, and GSDs, as well as to genotypic specificities, such as the number of spikes per SU (spike overlap), spike distribution, color, inclination, shape, size, presence of awns, and developmental stage. The canopy at the growth stage of spike development is very complex, with variations in its appearance among genotypes, herein also specified in detail. A technical, as well as the timely, easy adaptation of a methodology to the aforementioned external conditions and genotypic specificities would be beneficial and remains a significant challenge.

In the present study, UAV-based RGB images and the color space transformation in the extended CIELAB color space were used to determine the wheat spike cover (number of spike pixels per square meter). Influenced by the homogeneous appearance of spikes in the canopies, a linear relation between the spike cover and the number of spikes can be expected. Using this hypothesis, empirical regression models were trained for different wheat varieties, based on manually counted spikes and corresponding spike cover. The approach makes use of a simple color space transformation (extended CIELAB) for spike cover determination and does not require expert knowledge, labelled data or large training datasets to identify, for instance, overlapping and intersecting spikes. This potential enables a simple adaptation to the external conditions of the scene and genotypic specificities, by means of a recalibration of the model.

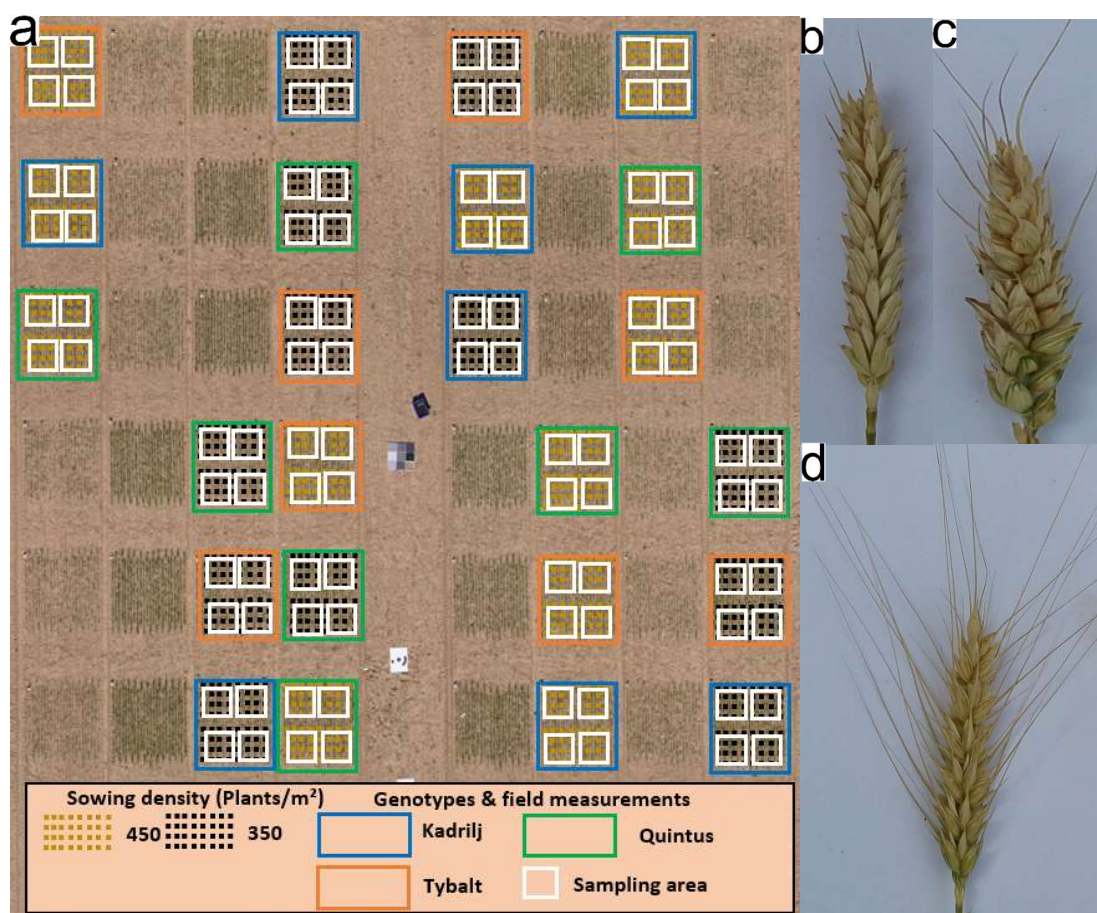
The presented study aims to identify factors influencing the accuracy of the developed empirical regression models to quantify spikes from UAV data, and tries to answer the following questions: (i) What is the ideal wheat developmental stage and time frame for data acquisition? (ii) What influence does the GSD have on regression model accuracies? (iii) What impact do spike and canopy characteristics (e.g., genotype, presence of awns, spike inclination) have on regression model accuracies? (iv) How many reference measurements are necessary for a user-based calibration of an empirical regression model, adapted to the environment of the scene?

## 5.2 Material and methods

### 5.2.1 Study site and experimental design

Three different spring wheat genotypes (Tybalt, Kadrijl and Quintus) were arranged in a randomized block design in the form of small breeder plots (1.4×3 m) at the agricultural research station, Campus Klein-Altendorf. The genotypes were sown in rows with a space of 10.4 cm and a density of 350 and 450 seeds m<sup>-2</sup>, with four repetitions for each sowing density.

In addition, each plot was divided into four individual SUs of one square meter (Figure 22a), resulting in 32 samples per genotype for further analysis. The field, with a total area of 900 m<sup>-2</sup>, was treated with 80 kg N/ha and growth inhibitors.



**Figure 22** Experimental layout, with highlighted sampling areas and genotypic specificities for method development. (a) Experimental field consisting of small breeder plots, with two sowing densities and three genotypes; (b) Tybalt with an awnless spike development; (c) Kadrijl with a low to medium rate of awn occurrence, and (d) Quintus with a high rate of awn occurrence.

Selected wheat genotypes varied in terms of spike and awn characteristics. Tybalt has an awnless spike development (Figure 22b), whereas Kadrijl and Quintus can be characterized by a low to a high rate of awn occurrence (Figure 22c-d). This variation in sowing densities,



spikes and awn characteristics' potential enable the robustness and the repeatability of the approach to be evaluated.

### 5.2.2 UAV data acquisition

For UAV data acquisition, the Sony Alpha 6000 RGB camera was used, equipped with a fixed focal length lens of 30 mm. Light sensitivity was set to ISO 800 and the white balance was fixed to the daylight setting for equal image acquisition settings.

Over a period of one month, the experiment was repeated four times (Table 13). Data were acquired at a wind speed of less than 15 km/h and in overcast conditions.

**Table 13** Information on UAV acquisition dates, growth stage of the spring wheat according to the BBCH scale and approximate wind speed during data acquisition.

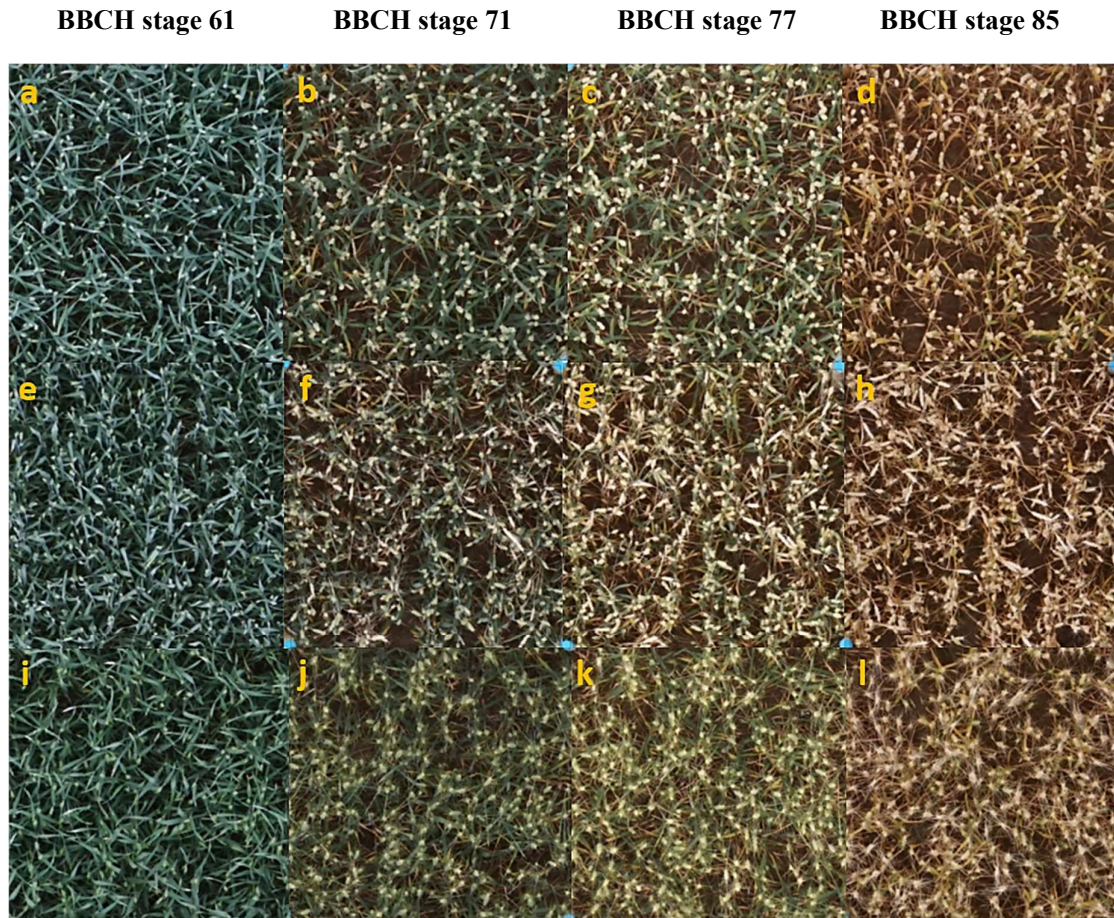
Dataset	Date	BBCH stage	Wind speed (km/h)
1	14.06.2018	61	< 10
2	05.07.2018	71	< 10
3	10.07.2018	77	< 15
4	18.07.2018	85	< 15

According to the BBCH scale, the developmental stages varied between early flowering (BBCH 61) and ripening (BBCH 85). The fruit development (BBCH 71 and 77) was expected to be the most appropriate growth stage for spike identification and was covered twice (Table 13). The properties of the different genotypes clearly affected the canopy characteristics at different developmental stages, as shown in Figure 23.

Image data were recorded from three different flight altitudes, at 8 m, 12 m and 17 m AGL for each BBCH stage, to investigate the impact of a decreasing spatial resolution on the accuracy of spike identification. The corresponding GSDs at different flight altitudes, the recorded number of images and the flight durations are summarized in Table 14. If the GSD was reduced, for example, from 1.4 mm to 2.0 mm, the number of acquired images and the flight duration to cover the region of interest was reduced by more than the half.

**Table 14** Overview of corresponding GSDs, number of images and flight duration to cover the region of interest, depending on the flight altitude.

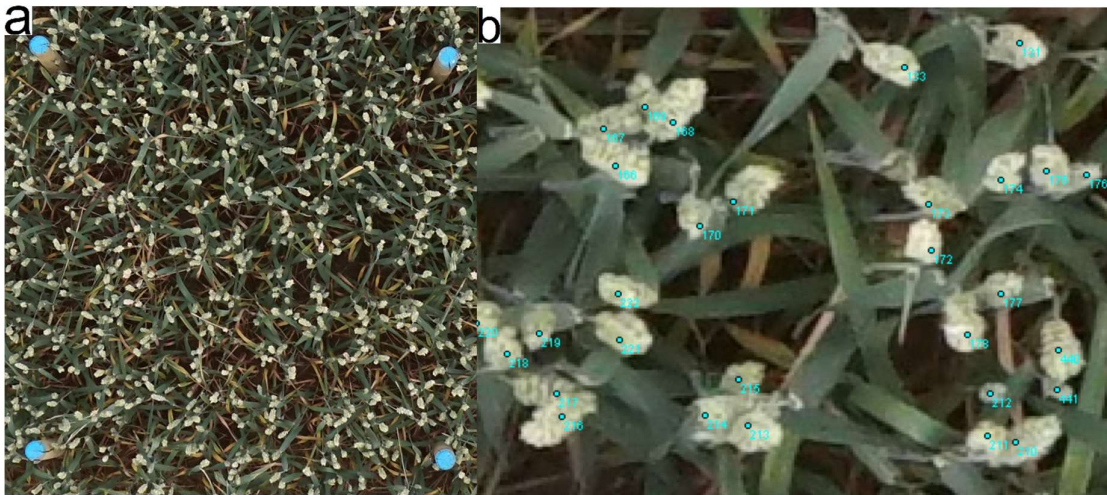
Flight altitude (m)	GSD (mm)	Number of acquired images	Flight time (minutes)
8	1.4	460	~ 12
12	2.0	209	~ 5
17	2.8	113	~ 3



**Figure 23** UAV images from 8 m above ground level, illustrating the different canopy characteristics and colors of (a-d) Tybalt; (e-h) Kadrijl, and (i-l) Quintus at different BBCH stages from fruit development (BBCH 61) to ripening (BBCH 85) in identical areas.

### 5.2.3 Acquisition of the training dataset

The spike density of each SU was manually counted using the software ImageJ (Schneider et al. 2012) from the UAV images, with a GSD of 1.4 mm, recorded at BBCH stage 71. The corners of the SUs were marked with blue wooden stakes in the field, to identify the region of interest within each image (Figure 24a). Each time images were chosen, the SU was located in their central region. Due to the high spatial resolution and the selected developmental stage, the individual spikes could easily be counted (Figure 24b). The advantage of BBCH stage 71, as reference point for manual counting, is that spikes located deeper in the canopy, as well as potentially green spikes with delayed development, could be clearly identified.



**Figure 24** (a) Section of a UAV image at BBCH stage 71, used for manual counting with blue wooden stakes, facilitating the identification of the SU and (b) Enlarged section of a UAV image, with manually labelled spikes in the software ImageJ.

### 5.2.4 Spike segmentation in the extended CIELAB color space

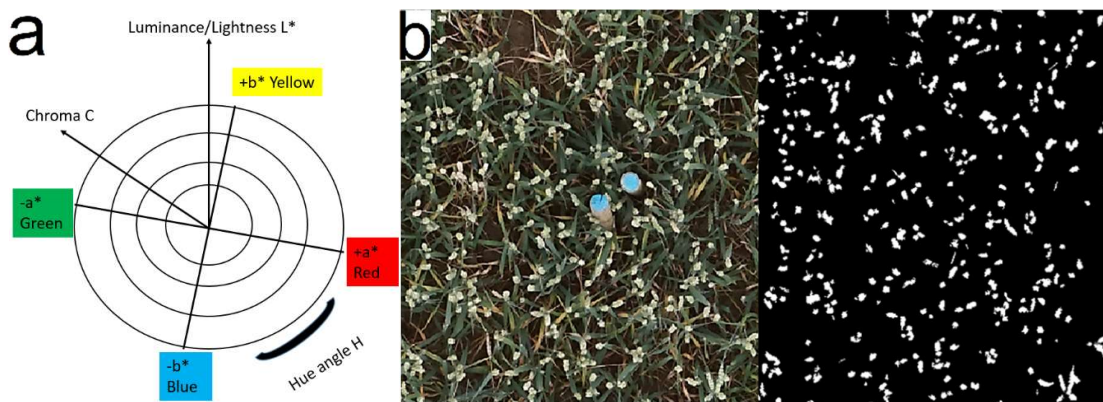
The CIELAB color space consists of a plane, spanned by the opponent-color dimensions green-red ( $a^*$ ), blue-yellow ( $b^*$ ), and the orthogonal luminance dimension ( $L^*$ ) (Figure 25a). This color space, with nonlinear, opponent-color dimensions was developed to simulate the response of the human eye. The transformation of RGB data into the CIELAB color space is an established method, aimed at distinguishing target objects, such as diseases, weeds or plants from the background (Kulkarni and Patil 2012; Ferreira et al. 2017; García-Martínez et al. 2020; Riehle et al. 2020). The color space can also make use of two additional dimensions, the chroma (C) and the hue angle (H) (Figure 25a). In particular, the C dimension is valuable in terms of coping with the high variation of green colors, e.g., in plant canopies during spike development (Figure 23). Details of the RGB transformation to the extended CIELAB color space can be found in several publications (Kurniawati et al. 2009; Macedo-Cruz et al. 2011; Choudhary et al. 2012; Azetsu and Suetake 2019).



In this study, these different dimensions were used to compute a binary image for spike segmentation. Values between the respective upper and lower thresholds were assigned to the spikes and acquired the value of one in the binary channel (Equation 6). The binary outcome was calculated by applying lower and upper thresholds in each dimension separately and then combining individual segmentation results. At the end, only those pixels in the binary image were identified as spikes, which have values of one in all channels, as illustrated by Equation 6. Otherwise, they are classified as zero (background).

$$Binary\ image(x, y) = \begin{cases} 1 & \text{if } \begin{cases} L^*_{low} < I_{L^*}(x, y) < L^*_{upp} \\ a^*_{low} < I_{a^*}(x, y) < a^*_{upp} \\ b^*_{low} < I_{b^*}(x, y) < b^*_{upp} \\ C_{low} < I_C(x, y) < C_{upp} \\ H_{low} < I_H(x, y) < H_{upp} \end{cases} \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

The necessary dimensions and thresholds for accurate spike segmentation, using the extended CIELAB color space, are hereinafter specified in detail. MATLAB environment 9.2 was used for image transformation, as well as for the computing of binary images. For a robust spike identification, the luminance dimension ( $L^*$ ) was well-suited (Figure 25b). The removal of the green, as well as the yellow parts of the canopy, with a luminance similar to the spikes, was achieved using the  $C$  dimension. In the early stages of the development, the lower threshold in the  $a^*$  dimension was also suitable for this. Furthermore, the upper threshold in the  $a^*$  dimension was ideal for removing the brown areas of the canopy, with a luminance similar to the spikes. Dimension  $H$  was primarily used to remove the blue wooden stakes (Figure 25b), as well as a number of very light brown areas within the canopy.



**Figure 25** (a) Dimensions of the extended CIELAB color space that was utilized for spike segmentation; (b) Left: section of the RGB image; Right: corresponding binary image, segmented in the extended CIELAB color space. White pixels represent the spikes and black pixels represent the background.

The lower and upper thresholds in the extended CIELAB color space were identified separately for each BBCH stage, GSD, and genotype to ensure an optimal spike segmentation. The applied thresholds are summarized in Table S5.

In addition, two filters, based on fragment removal and hole-filling, were applied to the resulting binary images (Figure 25b). Firstly, all fragments, i.e., small, connected components, classified as spikes, were classified as background, if they only contained 10 pixels or less. Secondly, all holes, i.e., small, connected components, classified as background, were treated in an analogous manner.

Finally, shapefiles, based on the aforementioned highlighted area, were used to calculate the spike cover (number of spike pixels per square meter) of the individual SUs in the processed orthomosaics (Figure 22a).

### **5.2.5 Statistical analysis**

The manually counted spike number in combination with the determined spike cover per SU enabled the build-up of empirical regression models for spike density assessment. During this process, the data containing 32 samples for each genotype were randomly divided 16 times into a calibration and validation set. Two thirds of the samples ( $n = 21$ ) were used to train the model (calibration), while the remaining third ( $n = 11$ ) was used to evaluate the model accuracy (validation). The performance of all models was determined based on the  $R^2$ , RMSE, and the relative root mean square error (rRMSE) using the software R (2018). In order to identify a significant difference between the models, two-way ANOVAs were conducted to verify the RMSE, including a Tukey post-hoc test.

Furthermore, the study makes use of the power analysis to estimate an appropriate training sample size for the best-performing models of each genotype and GSD. The power analysis is a test of probability to correctly reject the null hypotheses (Lerman 1996; Tomczak et al. 2014). In this study, the null hypothesis assumes no significant difference in prediction accuracy for an appropriate training sample size ( $p = 0.05$ ). A minimum power of 0.8 was set by the statistical literature to accept the null hypothesis (Cohen 1988; Lerman 1996; Tomczak et al. 2014). In addition, the datasets were again randomly divided 16 times into a calibration and a validation set, whereby the training sample size was reduced stepwise. Each time, either 16, 11 or 6 samples were used for model calibration. The remainder of the total sample sizes enabled the validation of the model performance, based on the averaged  $R^2$ , RMSE, and rRMSE. In summary, the investigations attempted to estimate the effort required to adapt the model to specific scene conditions.

## 5.3 Results

### 5.3.1 Tybalt

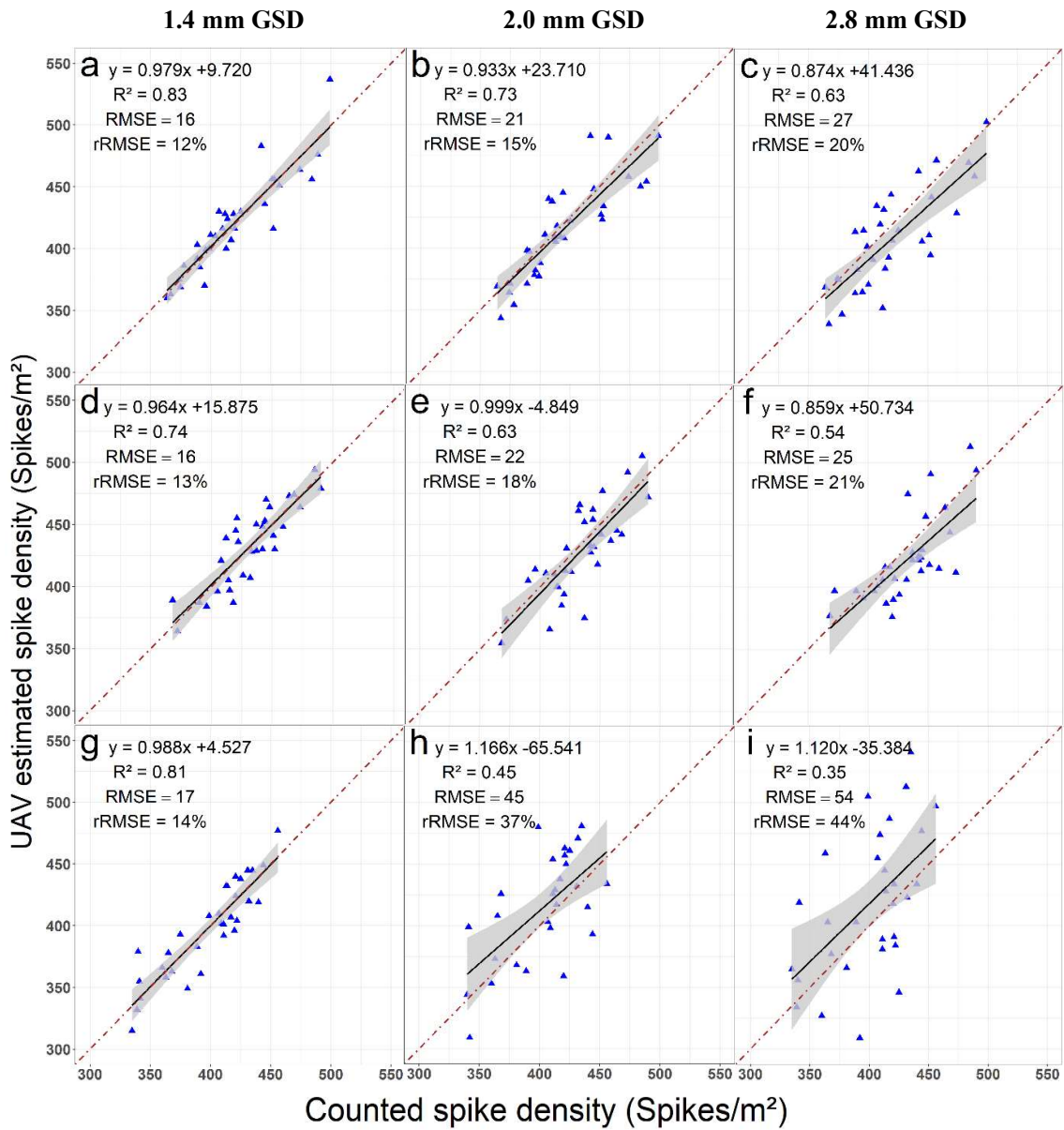
The results of the models, predicting the spike density of Tybalt at different developmental stages and GSDs, are summarized in Table 15. At the beginning of the observation period, during flowering at BBCH stage 61 (Figure 23a), a minimal contrast between the spikes and the other canopy elements can be observed. The leaves frequently had a similar luminance and color, and it was difficult to discern spikes alone. Model development, based on data with the highest GSD, only led to a low  $R^2$  of 0.47 and a high RMSE of 40 spikes  $m^{-2}$ . Subsequently, lower resolved image data, acquired in the same developmental stage, was not suitable for a meaningful model development (Table 15).

**Table 15** Statistical relationship between the manually counted spike density and spike cover of Tybalt at different BBCH stages and GSDs. Averaged statistical measures were computed across the empirical regression models, randomly divided into calibration and validation datasets. Different letters indicate significant differences between the RMSE ( $p < 0.05$ ).  $R^2$ : averaged coefficient of determination; RMSE: averaged root mean square error; rRMSE: relative root mean square error.

BBCH	GSD (mm)	Calibration (n = 21)			Validation (n = 11)				
		Slope	Intercept	$R^2$	Slope	Intercept	$R^2$	RMSE	rRMSE (%)
61	1.4	0.018	-2.784	0.49	1.015	-18.040	0.47	40g	30
	2.0	0.002	3.449	0.02					
	2.8	0.000	1.173	0.01			/		
71	1.4	0.021	-0.475	0.82	0.979	9.720	0.83	16a	12
	2.0	0.021	1.920	0.73	0.933	23.710	0.73	21bc	15
	2.8	0.016	1.941	0.65	0.874	41.436	0.63	27ef	20
77	1.4	0.025	-2.462	0.80	1.002	-0.176	0.80	18ab	13
	2.0	0.018	1.194	0.68	0.947	12.999	0.71	24cd	17
	2.8	0.018	1.199	0.58	0.940	13.744	0.57	32f	23
85	1.4	0.019	0.627	0.70	0.980	5.996	0.68	24de	18
	2.0	0.019	0.326	0.64	1.109	-54.423	0.64	32f	24
	2.8	0.016	1.705	0.49	0.964	20.160	0.51	34f	25

During fruit development (BBCH 71), the contrast between the spikes and the other canopy elements was more clearly visible: the spikes had a light grey color, whereas the rest of the canopy was predominantly dark green (Figure 23b,c). Therefore, the best results for the different GSDs were obtained in this developmental stage (Table 15). The dataset with the highest GSD of 1.4 mm led to a model having an rRMSE of 12 % and an RMSE of 16 spikes  $m^{-2}$ . The model also showed less scattering and the regression line indicated a good match to the 1:1 line (Figure 26a).





**Figure 26** Statistical relationship between the manually counted spikes and from the empirical regression model estimated spike density of the three GSDs and the best-performing models of each genotype; **(a-c)** Tybalt at BBCH stage 71; **(d-f)** Kadrij at BBCH stage 77, and **(g-i)** Quintus at BBCH stage 85. Averaged statistical measures were computed randomly across the division of the empirical regression models. Black lines represent regression lines with 95 % confidence intervals; brown lines represent the 1:1 line.

With a lower GSD (2.0 mm), a slight underestimation of the predicted spike density was observed (Figure 26b) and the RMSE increased to 21 spikes m<sup>-2</sup>. For the dataset with the lowest spatial resolution (GSD = 2.8 mm), the underestimation of the predicted spike density identified, was accompanied by a higher scattering and a lower model performance ( $R^2 = 0.63$ , RMSE = 27 m<sup>-2</sup>, rRMSE = 20 %) (Figure 26c).

During late fruit development (BBCH 77), the RMSE of the models increased slightly by a maximum of five spikes  $m^{-2}$ , compared to BBCH stage 71 (Table 15).

By contrast, the model performances were significantly reduced in the subsequent growth stage at BBCH stage 85. Spikes and other canopy elements matured and turned yellow, which led to a reduction in contrast between the spikes and the rest of the canopy (Figure 23d). The model, based on the images with the highest GSD, still had an  $R^2$  of 0.68 with an RMSE of 18 spikes  $m^{-2}$  (Table 15). With a lower spatial resolution of the data (GSD > 2.0 mm), the RMSE was higher than 24 spikes  $m^{-2}$ .

### 5.3.2 Kadrilj

The model performances of Kadrilj during flowering (BBCH 61) were also negatively affected by the weak contrast of the spikes and other canopy elements (Figure 23e). This was additionally amplified due to the awn occurrence of the genotype. Model development was only meaningful, if at all, based on the data with highest GSD, which still resulted in low prediction accuracy with an  $R^2$  of 0.21 and an RMSE of 67 spikes  $m^{-2}$  (Table 16).

**Table 16** Statistical relationship between the manually counted spike density and spike cover of Kadrilj at different BBCH stages and GSDs. Averaged statistical measures were computed across the empirical regression models, randomly divided into calibration and validation datasets. Different letters indicate significant differences between the RMSE ( $p < 0.05$ ).

BBCH	GSD (mm)	Calibration (n = 21)			Validation (n = 11)				
		Slope	Intercept	$R^2$	Slope	Intercept	$R^2$	RMSE	rRMSE (%)
61	1.4	0.011	6.308	0.21	1.490	-97.227	0.21	67e	55
	2.0	0.003	2.528	0.02	/				
	2.8	0.001	1.648	0.01					
71	1.4	0.024	-2.490	0.78	1.059	-28.716	0.77	17ab	14
	2.0	0.022	-0.637	0.53	0.973	5.733	0.59	25c	20
	2.8	0.019	1.337	0.34	1.062	-29.853	0.33	44d	36
77	1.4	0.026	-2.393	0.76	0.964	15.875	0.74	16a	13
	2.0	0.026	-2.609	0.64	0.999	-4.849	0.63	22c	18
	2.8	0.027	-2.881	0.53	0.859	50.734	0.54	25c	21
85	1.4	0.006	7.751	0.04	/				
	2.0	0.005	3.618	0.03					
	2.8	0.006	7.497	0.02					

Similar to Tybalt, the contrast between the spikes and the other canopy elements was considerably enhanced during the stage of fruit development (Figure 23f,g). At BBCH stages 71 and 77, the highest model accuracies could be achieved, based on the data with a GSD of 1.4 mm (RMSE < 17 spikes  $m^{-2}$ , rRMSE < 14 %) (Table 16). The regression line also showed a good match to the 1:1 line, without over- or underestimation of the predicted spike density (Figure 26d).

The model, based on image data and having a GSD of 2.0 mm at BBCH stage 71, led to an RMSE of 25 spikes m<sup>-2</sup> (Table 16). In the case of images with the lowest spatial resolution, the model performance further declined (RMSE = 44 spikes m<sup>-2</sup>, rRMSE = 36 %). The awns of the genotype affected the color and luminance of the spikes. In comparison to Tybalt, the spikes were still partly green during early fruit development (Figure 23b,f). This negatively affected the segmentation and the model performances, particularly in the case of image data having the lowest GSD of 2.8 mm.

Influenced by the further development of the plants at BBCH stage 77, ideal model performances were finally observed for the image data of the two lowest GSDs (GSD = 2.0 mm, RMSE = 22 spikes m<sup>-2</sup>, rRMSE = 18 %; GSD = 2.8 mm, RMSE = 25 spikes m<sup>-2</sup>, rRMSE = 21 %). For corresponding models, an underestimation of higher values, in particular, were observed (Figure 26e,f).

At BBCH stage 85, a meaningful assessment of spike cover and correspondingly spike density, was not possible for Kadrijl (Table 16). The ripening clearly reduced the contrast between the spikes and the other canopy components (Figure 23h).

### 5.3.3 Quintus

The accuracy of the spike density assessment of Quintus was significantly dependent on the spatial resolution, as well as the developmental stage. During flowering (BBCH 61), the model development was independent of the GSD and was not meaningful (Table 17).

**Table 17** Statistical relationship between the manually counted spike density and spike cover of Quintus at different BBCH stages and GSDs. Averaged statistical measures were computed across the empirical regression models, randomly divided into calibration and validation datasets. Different letters indicate significant differences between the RMSE ( $p < 0.05$ ).

BBCH	GSD (mm)	Calibration (n = 21)			Validation (n = 11)				
		Slope	Intercept	R <sup>2</sup>	Slope	Intercept	R <sup>2</sup>	RMSE	rRMSE (%)
61	1.4	0.001	1.499	0.06	/				
	2.0	0.001	0.741	0.03					
	2.8	0.001	0.509	0.03					
71	1.4	0.016	1.943	0.52	0.956	29.934	0.54	32b	26
	2.0	0.013	3.232	0.19	0.812	54.793	0.17	65d	53
	2.8	0.009	5.236	0.12	0.722	130.955	0.09	81e	67
77	1.4	0.022	-0.398	0.55	0.969	16.788	0.56	31b	26
	2.0	0.017	1.729	0.30	0.998	1.859	0.28	55cd	45
	2.8	0.011	3.839	0.26	1.067	-30.754	0.25	63d	52
85	1.4	0.027	-3.390	0.81	0.988	4.527	0.81	17a	14
	2.0	0.029	-3.942	0.44	1.166	-65.541	0.45	45c	37
	2.8	0.032	-5.526	0.36	1.120	-35.384	0.35	54c	44

At BBCH stage 71, the model, based on image data with the highest GSD of 1.4 mm, led to an  $R^2$  of 0.54 and an RMSE of 32 spikes  $m^{-2}$  (Table 17). In the case of the models based on the lower resolved image data (GSD > 2.0mm), the correlation significantly declined, resulting in an RMSE of more than 65 spikes  $m^{-2}$  (Table 17). Due to the high rate of awn occurrence, the spikes at the stage of fruit development were still green and exhibited a slight contrast in comparison to the rest of the canopy (Figure 23j).

Image data, acquired at the later development stage of BBCH 77, led to a slightly increased model performance of the two lowest GSDs (GSD = 2.0 mm,  $R^2 = 0.28$ , RMSE = 55 spikes  $m^{-2}$ ; GSD = 2.8 mm,  $R^2 = 0.25$ , RMSE = 63 spikes  $m^{-2}$ ).

The model performances of Quintus were significantly improved in the subsequent growth stage (BBCH 85), when the spikes, the other canopy elements and, in particular, the awns were ripening (Figure 23i). The best model prediction accuracy was finally observed for the highest GSD, with an RMSE of 17 spikes  $m^{-2}$ , uncertainties of 14 % and an almost perfect match of the regression to the 1:1 line (Figure 26g). The reduction of the image GSDs (GSD > 2.0 mm) led to significantly lower model accuracies, with an  $R^2$  of less than 0.45 and an RMSE higher than 45 plants  $m^{-2}$  (Table 17). Additionally, the model showed a high level of scattering and an overestimation of predicted spike density, in particular, for lower values (Figure 26h,i).

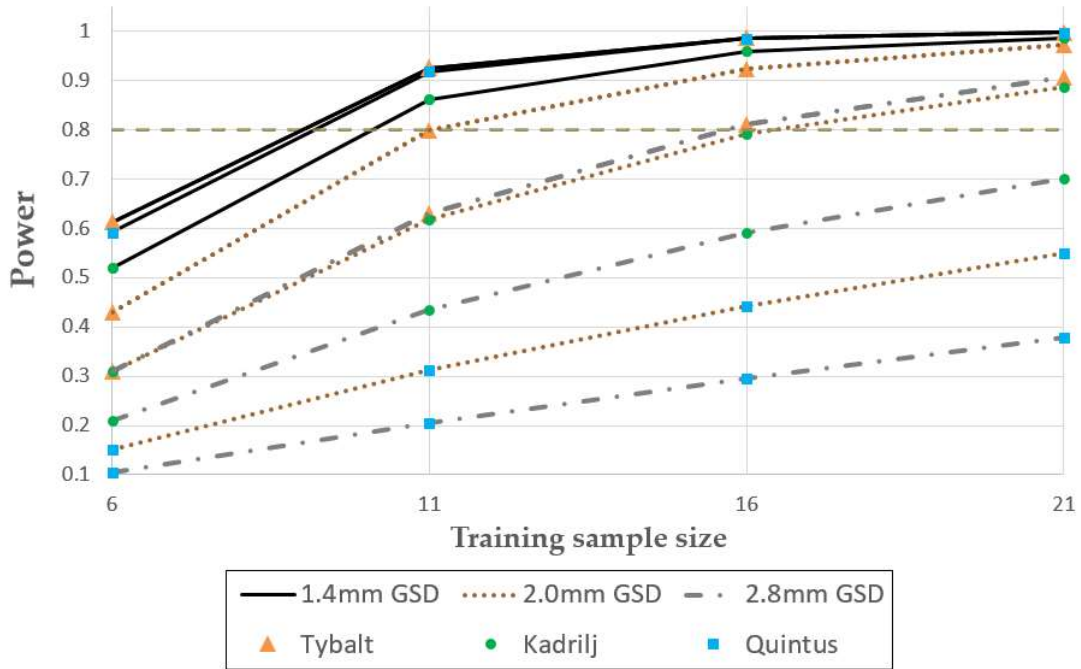
### **5.3.4 Power analysis and stepwise reduction of the sample size**

In order to identify the appropriate training sample size for a user-based calibration of an empirical regression model in this study, a statistical power analysis and a stepwise reduction of the training sample size were computed. The results are illustrated in Figure 27 and Table S6 for the best-performing models, developed for each genotype and GSD.

For the models with the highest GSD (1.4 mm) of the different genotypes, a continuous reduction of the training sample size from 21 to 11 samples still led to a power over 0.80 (Figure 27). When the training subset was reduced to six samples, the power clearly declined, with values between 0.61 and 0.52.

With a reduction in GSD to 2.0 mm, similar results were observed for Tybalt with a power over 0.80 for models calibrated with more than 11 samples (Figure 27). The investigation of the stepwise reduction in the training sample size also indicated that only a statistical power below 0.80 resulted in significant differences in model performances in this study (Table S6, Figure 27). Therefore, for Kadrijl and the aforementioned GSD of 2.0 mm, the first reduction step, with a power less than 0.80, also led to a significant difference in model performance (Table S6). However, the reduction of the training sample size from 21 to 11 samples increased the RMSE by a maximum of six spikes  $m^{-2}$  (Table S6). Similar results were observed for Tybalt, in combination with the model performance having the lowest GSD of 2.8 mm.

Due to the poor model accuracies achieved for Quintus, based on the images with GSDs lower than 1.4 mm and a determined power of less than 0.55 for the highest number of training samples (Figure 27), a reduction of the training subset was not meaningful. The stepwise reduction of the training sample size also identified significantly decreasing model accuracies (Table S6).



**Figure 27** Power curves, based on the number of samples determined for the best-performing models of Tybalt at BBCH stage 71, Kadrij at BBCH stage 77, and Quintus at BBCH stage 85, to estimate the appropriate training sample size. An averaged statistical power analysis was computed across the empirical regression models, randomly divided into calibration and validation datasets ( $p = 0.05$ ).

## 5.4 Discussion

The results of the study demonstrated that the spike density of differential wheat phenotypes can be predicted with uncertainties of less than 14 % from auto-captured UAV images for high-throughput field phenotyping.

In the case of Tybalt, which is an awnless genotype, the early fruit development stage (BBCH 71) was identified as the most suitable growth stage for spike density assessment (Table 15). Models, based on image data acquired during late fruit development (BBCH 77), also provided comparable model performances with possible RMSEs of 18 spikes  $m^{-2}$ . The high contrast between spikes and other canopy elements enabled high prediction accuracies. In addition, the spikes of Tybalt were homogeneously distributed at a similar height level within the canopy, which was advantageous for their detection (Figure 23b,c). In the subsequent development stage 85, which corresponds to ripening, the contrast between the spikes and the other canopy elements was severely reduced. This led to a less precise segmentation and lower prediction accuracies (Table 15).

Moreover, the models of Tybalt, built on images with a spatial resolution of 2.0 mm during fruit development (BBCH 71-77) were able to predict spike density, with slightly lower accuracies (RMSE < 24 spikes  $m^{-2}$ ).

Similar to Tybalt, the fruit development of Kadrij (BBCH 71-77) was the ideal stage for spike density quantification. At this growth stage, the RMSE of the validation models was below 17 and 24 spikes  $m^{-2}$  for image data, with a GSD of 1.4 mm and 2.0 mm, respectively. In comparison to Tybalt, the model accuracy was slightly lower. This was due to a more heterogeneous canopy structure, characterized by spikes growing at different heights within the canopy and a more pronounced spike inclination (Figure 23f,g). Furthermore, Kadrij, with a low to medium rate of awn occurrence led to a greater color variation between spikes, in particular, at the early fruit development stage (Figure 23f). Over the course of further plant development, the number of previously dark green spikes became fewer, with potentially improved segmentation accuracy. Model development, based on image data with the lowest GSD (2.8 mm) thus led to better accuracy for late fruit development (BBCH 77).

As regards genotypes with a high rate of awn occurrence, such as Quintus, the most suitable growth stage for spike detection occurred relatively late in plant development (BBCH 85). The models, based on image data recorded during the earlier growth stages, were clearly influenced by the high rate of awn occurrence, leading to a lower contrast between the spikes and the other canopy elements (Figure 23j,k). The awns remained green until the ripening stage, which hindered the determination of sound thresholds. In addition, for good model performances, data acquisition was only meaningful in the case of images with the highest GSD of 1.4 mm (Table 17).

In summary, the spike density could be accurately modelled with an RMSE of less than 17 spikes  $m^{-2}$  among the investigated wheat phenotypes and ideal models. This performance is comparable to the results of other studies (Cointault et al. 2008; Fernandez-Gallego et al.

2019; Madec et al. 2019; Sadeghi-Tehran et al. 2019; Fernandez-Gallego et al. 2020). Moreover, the presented methodology is firstly advantageous in that the approach directly quantifies the spike density in standard units ( $\text{m}^{-2}$ ), compared to other studies that still require a selection of representative subsample areas (Cointault et al. 2008; Zhu et al. 2016; Alharbi et al. 2018; Fernandez-Gallego et al. 2018; Zhou et al. 2018a; Fernandez-Gallego et al. 2019). This needs an additional extrapolation step to obtain comparable standard units, leading to unknown uncertainties, especially in terms of a spatial variability in the distribution of spikes. Secondly, the results in this work suggest that the model accuracies did not degrade with an increase in spike density at the ideal GSD and BBCH stage (Figure 26a,d,g). Thirdly, images were auto captured at regular intervals from the UAV, according to a flight pattern. Therefore, it was not necessary to manually acquire images of each SU. Fourthly, a georeferenced orthomosaic was used for data analysis, instead of single images of each SU, ideally for large scales. Lastly, spike quantification was possible with a lower spatial resolution, compared to approaches applied to highly-resolved images ( $\text{GSD} < 0.60$  mm) (Cointault et al. 2008; Alharbi et al. 2018; Hasan et al. 2018; Zhou et al. 2018a; Madec et al. 2019; Sadeghi-Tehran et al. 2019). In this study, the spike density could be accurately modelled with a GSD of 2.0 mm for Tybalt and Kadrij (RMSE  $< 22$  spikes  $\text{m}^{-2}$ ). The reduction of the GSD from 1.4 mm to 2.0 mm already reduced the necessary time for image acquisition and the number of images to cover the region of interest by more than a half (Table 14).

The significantly declining model accuracies for Quintus, with a GSD lower than 1.4 mm (Table 17) can be associated with this reduced number of acquired images and the late developmental stage (BBCH 85). Due to larger image footprints at higher flight altitudes, the central location of an SU within the recorded images, classically used for creating the orthomosaic, was no longer ensured. This could potentially lead to errors arising from distortion or perspective (Jaud et al. 2019). In addition, the spikes during ripening (BBCH 85) were often inclined (Figure 231). Inclined spikes that were not reconstructed based on a central area of the image can be an error source for the determination of fractional cover. Thus, it can be assumed that this was the reason why the models, based on a lower GSD (2.0 and 2.8 mm) exhibited such a high scattering and essentially overestimated the spike density of Quintus (Figure 26h,i). For Tybalt and Kadrij, such distortion and perspective problems were less relevant, since the ideal models for spike assessment occurred during the earlier stages of development, without problematical spike inclination.

Therefore, it is recommended, particularly in the case of predominantly inclined spikes, that the SU be recorded by a central area of the image for a sufficient reconstruction within an orthomosaic. However, compared to Fernandez-Gallego et al. (2020), the results in this study showed that a manual UAV image acquisition is still not necessary to record this plot centroid. The set-up, based on a 1.4 mm GSD and UAV auto-captured image within the lawnmower pattern (60 % across and 80 % along the track overlap) ensured a high prediction accuracy and centrally located SUs in the acquired images, to reconstruct the orthomosaic sufficiently.



In this study, the influence of three distinguished wheat phenotypes and canopy characteristics on the method accuracy for spike density assessment was examined. The complex structure of the canopy and the variation among genotypes and stages are illustrated in Figure 23. Machine learning approaches (Zhu et al. 2016; Hasan et al. 2018; Zhou et al. 2018b; Madec et al. 2019; Sadeghi-Tehran et al. 2019), image processing techniques (Cointault et al. 2008; Liu et al. 2014; Alharbi et al. 2018; Fernandez-Gallego et al. 2018; Zhou et al. 2018a), and also the developed models in this study have been adapted to these genotypic specificities (spike overlap, distribution, color, inclination, shape, size, presence of awns, developmental stage) as well as to external conditions (illumination, background, soil moisture content, sensors). Therefore, methodologies may fail when they are applied to images acquired in relation to different external conditions and genotypic specificities. In the case of machine learning approaches, diversity is critical, thus training datasets from thousands to millions are required to show great advantages (Deng et al. 2009; Ubbens et al. 2018; Sadeghi-Tehran et al. 2019). In this study, the results led to the assumption that among the examined wheat phenotypes, 11 reference measurements were appropriate for a user-based calibration of an empirical regression model (Figure 27, Table S6). However, further investigations are necessary to evaluate the robustness of the adaptation process to different conditions.

In summary, the presented results are very promising, but further investigations into similar and different phenotypes are essential to evaluate the robustness of the statistical relationship and the reliability of spike density prediction. In addition, applied thresholds must be potentially adapted to external conditions or different phenotypes, similar to other studies with the need of defining specific thresholds for spike identification (Cointault et al. 2008; Zhu et al. 2016; Alharbi et al. 2018; Fernandez-Gallego et al. 2018; Zhou et al. 2018b; Madec et al. 2019).

Furthermore, the developed approach is dependent upon specific weather conditions, which restrict the practicability of the approach. Firstly, it is recommended that the wind speed should not be higher than 15 km/h, similar to this study (Table 13). This avoids issues in relation to image processing, caused by excessive spike movements during data acquisition. Secondly, data acquisition is limited to overcast conditions. Constant solar illumination conditions during data acquisition clearly reduced the contrast between the spikes and the other canopy elements, as illustrated in Figure S1. Similar results were observed by Fernandez-Gallego et al. (2020), whose images were also acquired from UAVs. With regard to direct illumination, a higher resolution appears to be required to distinguish the spikes from the other canopy elements.

## **5.5 Conclusions**

This study presented an easy to implement and low-cost approach for a non-invasive assessment of spike density, based on the statistical relationship between counted spike density and fractional cover. The determination of spike cover per unit of ground area takes advantage of the fact that under ideal conditions, the average spike size is relatively stable. For the first time, image acquisition and data analysis make use of auto-captured UAV imagery and processed orthomosaics, ideally for large scales. The influence of different wheat phenotypes, as well as the potential errors of distortion and perspective with a decreasing GSD, were also considered in the methodology under in-field conditions. Moreover, the necessary effort was made to investigate whether the methodology could be adapted to external conditions or genotypic specificities. The results led to the assumption that 11 reference measurements were appropriate for a user-based adaptation process. However, the robustness of such an adaptation process, as well as the statistical relationship, needs to be examined in further studies.

The quantification of spike density can already be used to identify yield potential in the early phases for breeding programs and for variety selections. Considering the variation of spike density in-field, caused by different nutrition and soil conditions, this can increase a farmer's yield estimations in real time. These reliable yield estimations can reduce costs by improving farming management, such as the transportation, harvesting and storage of cereals.

## 6. Synthesis

The research in this thesis was focused on methodologies to quantify vegetation traits with a high degree of accuracy and practicability. For traits like plant and spike density, high spatial and temporal resolutions were necessary for method development and evaluation. As has been shown, the acquired UAV data met all these criteria and could be used to estimate vegetation traits non-invasively for small and numerous plots. The applied approaches can be used as a standard protocol for trait quantification and can replace subjective and expensive manual labor, with the added benefit of providing reliable identification of desired varieties. Application of the developed methodologies necessitates limited prior knowledge, compared to SVM and CNNs. Moreover, the methods were able to increase the throughput of trait assessment and were generally not limited to single developmental stages, an important factor in terms of practicability. In this thesis, individual images covering only a small area of interest were replaced by georeferenced mosaicked images for trait extraction, which are mandatory, especially for large scales or site-specific management. For parametric regression methods, the investigated model adaptation, with a small number of reference measurements, can potentially facilitate adaptation to specific scene conditions. Moreover, the 3D information used for lodging assessment is almost independent of abiotic or external factors and does not require recalibration. These elements are crucial for application of the developed methods in practice and for reducing the hurdle of implementation in the decision-making process of PA and breeding. However, there are also limitations to consider, which are documented in the following sections.

### 6.1 Thresholding

Threshold determination influences the accuracy and applicability of an approach. Therefore, ideally, a threshold should be defined automatically and objectively.

The MAXCH was the basis for determining the threshold values used for quantification of the lodging percentage and severity. However, this approach depends on a homogeneity of the MAXCH and can cause problems through variations among genotypes and locations in the field. Moreover, the accuracy of the height extraction method is in the range of approximately  $\pm 5$  cm and can also vary for different areas in the field, particularly for large scales. In summary, although the threshold is determined automatically and objectively, the user still needs to check heterogeneity of the MAXCH visually or with statistical approaches before applying the procedure. In the worst case, different MAXCH values for areas and genotypes need to be considered in the workflow.

The second part of the study focused on plant density assessment and made use of the established unsupervised Otsu thresholding (Otsu 1979) to calculate fractional cover (Section 4.2.4). This work and various other UAV studies have demonstrated reliable separation of vegetation from soil, based on variance between the two classes (Torres-Sánchez et al. 2015; López-Granados et al. 2016; Li et al. 2019; Marcial-Pablo et al. 2019).

However, Meyer and Neto (2008) and Kittler and Illingworth (1985) claimed that the approach can also lead to insufficient accuracy and that the performance of Otsu thresholding is dependent on the value distribution and should ideally be bimodal.

In this study, paths between plots were created in the experimental design (Figure 17, Section 4.2.1) to avoid negative effects of field measurements (e.g., plant counting) on further plant growth. Similar designs have been used in studies in which the Otsu thresholding has been successfully applied (Torres-Sánchez et al. 2015; López-Granados et al. 2016; Li et al. 2019; Marcial-Pablo et al. 2019). Such a design automatically results in a higher proportion of pure soil pixels compared to plant pixels. For orthomosaics, this enables reliable separation of vegetation from soil, based on variance between the two classes. Designs without paths between plots, which are common in traditional agriculture, have proportionally lower soil and higher plant content, which affects the value distribution. Therefore, for such agricultural fields, it is necessary to investigate whether the Otsu threshold method and the developed approach also provide good results. For designs similar to this study with paths between plots, such as breeding trials, the Otsu thresholding is an appropriate and proven method.

For spike segmentation, thresholds were determined subjectively in the extended CIELAB color space and required manual adjustment to suit the conditions of the scene. Machine learning approaches, as previously applied by Fernandez-Gallego et al. (2020), offer the potential to increase the objectivity and automation capacity of spike identification. However, machine learning algorithms also needs recalibration based on the scene context (e.g., genotypic specificities, external conditions, sensor). This recalibration requires expert guidance in feature extraction, is tedious (as the spikes are tiny) and needs numerous labeled images. The method used in this thesis suffers from a lack of automatability and objectivity but can be adapted with limited prior knowledge.

In summary, it is useful to check the results of the threshold procedure. On the one hand, this reduces the automation potential. On the other hand, the methodology allows for easy verification of the thresholding accuracy based on clear criteria, ideally to also determine the traits for unknown datasets and genotypes.

## **6.2 Georeferencing**

In this study, the location of points in the clouds was identified by GCPs in the field using high-accuracy RTK-GNSS measurements. Although this is a standard UAV data processing procedure, placement, measurement, and removal of GCPs in the field are labor-intensive tasks and therefore difficult to apply in practice (Song and Wang 2019; Teppati Losè et al. 2020). The location of points can be additionally determined by spatial coordinates of the camera, with substantially less effort. In particular, recently developed commercial UAVs such as the DJI Phantom 4 RTK system with a dual-frequency global navigation satellite system (GNSS) receiver simplify the georeferencing of photogrammetric UAV products. Different studies using on-board GNSS RTK for data processing have already shown similar georeferencing accuracy in relation to the horizontal position (x- and y-axis) to that achieved

by the GCP-based approach (Stott et al. 2020; Štroner et al. 2020; Urban et al. 2020). This is particularly beneficial for trait extractions based on 2D products, e.g., for plant and spike density in this study. Determination of the vertical position (z-axis) from an on-board GNSS RTK system is associated with larger inaccuracies in height estimations. This problem is mainly caused by incorrect estimation of interior orientation parameters and the bundle adjustment. However, this issue can be avoided by using a small number of GCPs (at least one or ideally four GCPs), distributed in the area of interest (Przybilla et al. 2020; Štroner et al. 2020). Therefore, the number of GCPs required can be reduced, making usage of UAVs more practicable for large-scale applications.

### **6.3 Reflectance calculation**

The DLS, in combination with a calibrated reference panel, was used in this study to radiometrically correct the multispectral images. This approach requires minimal time, resources, and human effort to estimate surface reflectance. In addition, this correction procedure enables irradiance variation to be addressed during data acquisition. However, different studies have indicated that the approach can also lead to an overestimation of reflectance (Cao et al. 2019; Doughty and Cavanaugh 2019). Moreover, Hakala et al. (2018) found that the vertical orientation of the irradiance sensor is important, and a tilt of only a few degrees can greatly affect the accuracy of the determined irradiance. The attitude of the UAV during a flight mission can easily change by  $\pm 5^\circ$  and was identified as a major error source (Hakala et al. 2018). In addition, MicaSense noted that the accuracy of measured irradiance is influenced by the sensor orientation relative to the sun's position (MicaSense 2017). In summary, the accuracy can be influenced by several factors, also leading to studies that have been observed good results in application and evaluation of the calibration process (Mamaghani et al. 2018; Su et al. 2019; Taddia et al. 2020).

In the present thesis, image data were acquired around solar noon to avoid issues caused by the position of the sun. In addition, the patented V-shaped UAV design allows the DLS to be installed on the gimbal. The gimbal, as an actively stabilized platform, specifically avoids the issue of irradiance measurements associated with a tilted DLS. However, this is only an exception, compared to commonly used UAV systems, as shown in Section 2.3.3. Sufficiently stabilized platforms for an upward-looking sensor need to be developed and cannot be installed on all UAV platforms, such as with fixed wings. Therefore, further investigations on novel and reliable tilt correction methods for on-board irradiance sensors, as already investigated by Suomalainen et al. (2018), have great potential and offer a more general solution independent of the UAV system used.

The empirical line method (ELM) described by Smith and Milton (1999) is an alternative approach widely applied to calculate estimates of surface reflectance from multispectral remote sensing data. It is based on the empirical relationship between the DN values of recorded image data and the at-surface reflectance of well-characterized reference targets. The determined prediction equations for each band take into account atmospheric

factors affecting the scene (Smith and Milton 1999). The ELM has a high degree of reliability in terms of image calibration and has already been applied and investigated in different studies (Baugh and Groeneveld 2008; Bondi et al. 2016). However, for the ELM approach, it is essential to purchase several reference targets (of sufficient quality and suitability), to place these in the field and measure the reflectance of targets at the scene with a spectrometer. In addition, it is necessary to go to the trouble of creating and applying the ELM as preprocessing for every image, a time-consuming and computationally demanding process. Furthermore, the ELM is not applicable under changing illumination conditions during data acquisition and is thus limited to cloudless conditions with constant illumination.

In summary, the ELM processing procedure and placement, measurement, and removal of targets in the field are labor-intensive. Moreover, the approach is only applicable under constant solar illumination conditions, which clearly limits the practicability of the ELM and highlights the advantage of an on-board irradiance sensor for radiometric correction of multispectral images.

## **6.4 Empirical-statistical regression methods**

In this thesis, parametric regression methods were used to assess plant and spike density, for the following reasons:

- 1) The required spatial resolution also allows use of UAVs for trait assessment, increasing the practicability of data acquisition and analysis;
- 2) These methods facilitate accurate and robust model performance;
- 3) A simple and small number of reference measurements facilitate investigation of the method accuracy for unknown datasets;
- 4) A small number of reference measurements potentially allow adaptation of the model to specific scene conditions;
- 5) Relatively little training and knowledge is required in order to apply and adjust the method;
- 6) Parametric regression methods are computationally fast.

However, the study results illustrated that, for plant density assessment and for model self-calibration, the timing of reference measurements (BBCH 11) differs from the timing of UAV data acquisition (BBCH 12-15). For spike density assessment, the timing of reference measurements (BBCH 71) may also differ from the ideal timing of UAV data acquisition (BBCH 85). Hence, it is necessary to check the BBCH stage regularly in order to ascertain the right time for reference measurement and UAV data acquisition. Influenced by ideal weather conditions, e.g., during plant germination in April in the study (Figure 3, Section 2.2), the time difference between BBCH stages was only four days (Table 8, Section 4.2.3). The need to measure twice at the right place and time for model self-calibration is clearly a drawback of the method. Identification of the right timing with the assistance for example of PhenoCam image data can help make this task more manageable.

Non-parametric regression methods such as SVM or CNNs are also dependent on the right timing for data acquisition and trait quantification (Zhou et al. 2018b; Jin et al. 2017b; Aich et al. 2018). In contrast, a single data set is sufficient for data labeling and trait extraction but, similarly, requires adaptation to specific situational conditions (e.g., genotypic specificities, sensor). This recalibration requires expert knowledge and needs numerous labeled images. Nevertheless, training datasets with a high degree of diversity have the potential to assess traits without the need for adaptation in the future. However, non-parametric regression methods (e.g., for plant density assessment) require much higher spatial resolution and are limited to early and specific developmental stages for trait extraction (Liu et al. 2017a; Jin et al. 2017b; Liu et al. 2017b). The developed parametric regression methods have higher variability in terms of developmental stage and GSD, with a positive impact on the practicability.

## 6.5 Hardware and software requirements

The developed methods are not computationally intensive and can be applied and adapted with limited prior knowledge. However, hardware and software requirements also influence the practicability of an approach. Table 18 summarizes the requirements for trait extraction in this thesis, including potential investment costs. The UAV price indicates the minimum price for a reliable and simple system that allows RGB waypoint data acquisition. UAV systems that allow simultaneous data acquisition with third-party multispectral and RGB cameras, as shown in Section 2.3.3, cost up to 6,000 € without batteries.

**Table 18** Hardware and software requirements, including investment costs for trait assessment in this thesis: X = mandatory; () = alternative

Equipment & Software	Data Acquisition	Sensor	Data Processing	Georeferencing		Others
	UAV & RGB	UAV & Multi- spectral	Photogramm- etric Software (e.g., Agisoft)	UAV RTK + Reference Station	RTK Device (GCPs)	PC & Drone License
Plant density	X	()	X	()	X	X
Spike density	X		X	()	X	X
Lodging	X		X		X	X
Investment costs (€)	> 1,500	~ 8,500	~ 3,000	~ 7,500	~ 5,000	~ 1,000

In summary, investment costs of at least 10.500 € are necessary for trait extraction (Table 18). In particular, georeferencing is costly. In addition to necessary basic knowledge about the UAV system and different steps (e.g., data processing and georeferencing), the requirements of the technology are likely to present some hurdles for a layperson. Moreover, based on the new EU regulations valid for 2021, a drone license is required even for small



UAVs with a take-off weight more than 250 g. These factors lead to the assumption that external companies may be better placed to implement UAV concepts in agriculture and farm management.

## 7. Conclusion and outlook

In this study, UAV image data were used to assess spatial variabilities in small-scale experimental plots. The final section of the thesis outlines the opportunities and major challenges for sustainable integration of such UAV applications in breeding and agriculture.

Growing attention is being paid in social sciences to the need for sustainable agriculture, and in recent decades, the EU has been keeping a close eye on environmental protection and nature conservation. Regulations and compliance will become more restrictive in the future so that negative environmental impacts can be reduced. Under these conditions, site-specific farm management will be essential in order to maintain steady gains in yields and to ensure global food security. Surveys of agricultural fields, as well as large breeding trials, with data of high spatial and temporal resolution will deliver important information for monitoring, as example, the effects of environmental conditions on crop phenology. Remote sensing with UAVs can provide powerful means in plant phenotyping and decision-making. Relevant data can be acquired in a short space of time, taking into account geolocations on a field scale, for further processing and site-specific recommendations for actions.

The key factor for meaningful implementation is the degree of heterogeneity within breeding trials or agricultural fields. The variability must at least be higher than the investigated regression error metrics (e.g., RMSE and MAE) of the developed methods. In addition, methodologies are usually specific to particular developmental stages, necessitating data acquisition at the right time and place. However, this is a challenge in breeding trials characterized by different developmental stages among genotypes and requiring a multitemporal data set for a single trait. Furthermore, the correct timing of data acquisition in relation to the developmental stage still needs to be assessed, which can be time-consuming for large breeding trials. For external companies, this process can only be economical if they obtain the information from local customers or via specific technologies (e.g., PhenoCams) in the field. Overall, it is ideal if the method works for different development stages, such as the plant density assessment in this study. This avoids the need for a multitemporal data set for breeding trials, as well as multiple identification of the right timing for data acquisition.

Moreover, empirical-statistical regression methods are adapted to specific phenotypes, layout, and illumination conditions. Therefore, to increase the application potential in practice, it is advantageous to know approximately how much effort will be required to adapt the methods to unknown data sets. However, this also requires detailed investigations in order to evaluate the robustness of the adaptation process in different conditions, similarly required in the scope of this thesis. One good result is not sufficient to confirm the universal applicability in practice.

These factors, along with the limitations addressed in Section 6, are the reasons why application of UAVs for plant trait extractions in agriculture is still at a low level outside the scientific community. At present, the reliability of empirical-statistical regression methods is not at the same level as reference measurement in the field. However, due to the time-consuming nature of manual surveys, only small potentially representative subareas are usually surveyed for extrapolation into standard units ( $m^{-2}$ ). Compared to UAV approaches, these extrapolations are subject to an unknown uncertainty, especially in the case of high spatial variability. UAV approaches can quantify plant traits directly in standard units with a known uncertainty. In addition, field measurements are not feasible for all applications, e.g., when a user needs timely information about a specific trait for decision-making and a manual survey would be too time-consuming and costly relative to the benefits. UAV approaches are quicker and less costly than manual surveys. They generate much more information, advantageous when it comes to optimizing the phenotyping process and for efficiency of management. Moreover, knowledge of geolocations allows development of on-board decision-making support for site-specific recommendations, such as fungicide application or acquisition of images with a higher resolution.

Field measurements that are limited to subjective estimations, such as the quantification of lodging, can easily be replaced by UAV applications. Moreover, procedures that make use of 3D information for trait assessment are much less dependent on the developmental stage or illumination conditions. They have high applicability and repeatability and may be an ideal substitute for manual surveys in the long term.

One can only guess which revolution will provide the solution to the problem of increased demand – superior genotypes or technological advances such as combination of surveying UAVs and small unmanned ground vehicles responsible for site-specific fertilization. It is highly likely that precision agriculture and remote sensing will contribute to this revolution. The major challenge in this regard will be to demonstrate universal applicability or satisfactory adaptation of the developed methods to meet the demands of precision agriculture and breeding.

## 8. Publications

1. Wilke, Norman, Bastian Siegmann, Lasse Klingbeil, Andreas Burkart, Thorsten Kraska, Onno Muller, Anna van Doorn, Sascha Heinemann, and Uwe Rascher. 2019. Quantifying lodging percentage and lodging severity using a UAV-based canopy height model combined with an objective threshold approach. *Remote Sensing* 11: 515. <https://doi.org/10.3390/rs11050515>

Contribution: Design 65 %, Experimentation 80 %, Analysis 85 %, Publication work 85%

2. Wilke, Norman, Bastian Siegmann, Johannes A. Postma, Vera Krieger, Onno Muller, Ralf Pude and Uwe Rascher. 2021. Assessment of plant density for barley and wheat using UAV multispectral imagery for high-throughput field phenotyping. *Computers and Electronics in Agriculture* 198. <https://doi.org/10.1016/j.compag.2021.106380>

Contribution: Design 85 %, Experimentation 90 %, Analysis 90 %, Publication work 90 %

3. Wilke, Norman, Bastian Siegmann, Mark Müller-Linow, Erekle Chakhvashvili and Uwe Rascher. In preparation. Assessment of cereal plant density using UAV multispectral imagery for high-throughput field phenotyping.

Contribution: Design 85 %, Experimentation 90 %, Analysis 80 %.

## 9. References

- Aasen, Helge, Andreas Burkart, Andreas Bolten, and Georg Bareth. 2015. Generating 3D hyperspectral information with lightweight UAV snapshot cameras for vegetation monitoring: From camera calibration to quality assurance. *ISPRS Journal of Photogrammetry and Remote Sensing*. <https://doi.org/10.1016/j.isprsjprs.2015.08.002>.
- Aasen, Helge, Eija Honkavaara, Arko Lucieer, and Pablo J. Zarco-Tejada. 2018. Quantitative remote sensing at ultra-high resolution with UAV spectroscopy: A review of sensor technology, measurement procedures, and data correction workflows. *Remote Sensing*. <https://doi.org/10.3390/rs10071091>.
- Abdulridha, Jaafar, Yiannis Ampatzidis, Sri Charan Kakarla, and Pamela Roberts. 2020. Detection of target spot and bacterial spot diseases in tomato using UAV-based and benchtop-based hyperspectral imaging techniques. *Precision Agriculture*. <https://doi.org/10.1007/s11119-019-09703-4>.
- Adamchuk, Viacheslav I, Richard B Ferguson, and Gary W Hergert. 2010. Soil Heterogeneity and Crop Growth. In *Precision Crop Protection - the Challenge and Use of Heterogeneity*, ed. Erich-Christian Oerke, Roland Gerhards, Gunter Menz, and Richard A. Sikora, 3–16. Dordrecht: Springer Netherlands. [https://doi.org/10.1007/978-90-481-9277-9\\_1](https://doi.org/10.1007/978-90-481-9277-9_1).
- Agisoft. 2018. Agisoft PhotoScan User Manual. *Professional Edition, Version 1.4*.
- Agisoft. 2019. MicaSense RedEdge MX processing workflow (including Reflectance Calibration) in Agisoft Metashape Professional 1.5 : Helpdesk Portal. Retrived from <https://agisoft.freshdesk.com/support/solutions/articles/31000148780-micasense-rededge-mx-processing-workflow-including-reflectance-calibration-in-agisoft-metashape-pro>, accessed January 02, 2022.
- Aich, Shubhra, Anique Josuttis, Ilya Ovsyannikov, Keegan Strueby, Imran Ahmed, Hema Sudhakar Duddu, Curtis Pozniak, Steve Shirliff, and Ian Stavness. 2018. DeepWheat: Estimating Phenotypic Traits from Crop Images with Deep Learning. In *Proceedings - 2018 IEEE Winter Conference on Applications of Computer Vision, WACV 2018*. <https://doi.org/10.1109/WACV.2018.00042>.
- Ainsworth, Elizabeth A., Claus Beier, Carlo Calfapietra, Reinhart Ceulemans, Mylene Durand-Tardif, Graham D. Farquhar, Douglas L. Godbold, et al. 2008. Next generation of elevated [CO<sub>2</sub>] experiments with crops: A critical investment for feeding the future world. *Plant, Cell and Environment*. <https://doi.org/10.1111/j.1365-3040.2008.01841.x>.
- Al-Mulla, Y. A., D. R. Huggins, and C. O. Stöckle. 2014. Modeling the emergence of winter wheat in response to soil temperature, water potential, and planting depth. *Transactions of the ASABE*. <https://doi.org/10.13031/trans.57.10200>.
- Alexandratos, N., and J. Bruinsma. 2012. World agriculture towards 2030/2050. *Land Use Policy*. [https://doi.org/10.1016/S0264-8377\(03\)00047-4](https://doi.org/10.1016/S0264-8377(03)00047-4).
- Alharbi, Najmah, Ji Zhou, and Wenjia Wang. 2018. Automatic counting of wheat spikes from wheat growth images. In *ICPRAM 2018 - Proceedings of the 7th International Conference on Pattern Recognition Applications and Methods*. <https://doi.org/10.5220/0006580403460355>.
- Anthony, David, Sebastian Elbaum, Aaron Lorenz, and Carrick Detweiler. 2014. On crop height estimation with UAVs. In *IEEE International Conference on Intelligent Robots and Systems*. <https://doi.org/10.1109/IROS.2014.6943245>.
- Araus, José Luis, and Jill E. Cairns. 2014. Field high-throughput phenotyping: The new crop breeding frontier. *Trends in Plant Science*. <https://doi.org/10.1016/j.tplants.2013.09.008>.

- Asseng, S., F. Ewert, P. Martre, R. P. Rötter, D. B. Lobell, D. Cammarano, B. A. Kimball, et al. 2015. Rising temperatures reduce global wheat production. *Nature Climate Change*. <https://doi.org/10.1038/nclimate2470>.
- Azetsu, Tadahiro, and Noriaki Suetake. 2019. Hue-preserving image enhancement in CIELAB color space considering color gamut. *Optical Review*. <https://doi.org/10.1007/s10043-019-00499-2>.
- Bareth, Georg, Helge Aasen, Juliane Bendig, Martin Leon Gnyp, Andreas Bolten, Andrés Jung, René Michels, and Jussi Soukkamäki. 2015. Low-weight and UAV-based Hyperspectral Full-frame Cameras for Monitoring Crops: Spectral Comparison with Portable Spectroradiometer Measurements. *Photogrammetrie - Fernerkundung - Geoinformation*. <https://doi.org/10.1127/pfg/2015/0256>.
- Bassine, Fatima Zahra, Ahmed Errami, and Mohammed Khaldoun. 2019. Vegetation Recognition Based on UAV Image Color Index. In *Proceedings - 2019 IEEE International Conference on Environment and Electrical Engineering and 2019 IEEE Industrial and Commercial Power Systems Europe, IEEEIC/I and CPS Europe 2019*. <https://doi.org/10.1109/IEEEIC.2019.8783830>.
- Basso, Bruno, Costanza Fiorentino, Davide Cammarano, and Urs Schulthess. 2016. Variable rate nitrogen fertilizer response in wheat using remote sensing. *Precision Agriculture*. <https://doi.org/10.1007/s11119-015-9414-9>.
- Baugh, W. M., and D. P. Groeneveld. 2008. Empirical proof of the empirical line. *International Journal of Remote Sensing*. <https://doi.org/10.1080/01431160701352162>.
- Bendig, Juliane, Andreas Bolten, and Georg Bareth. 2013. UAV-based Imaging for Multi-Temporal, very high Resolution Crop Surface Models to monitor Crop Growth Variability. *Photogrammetrie - Fernerkundung - Geoinformation* 6: 551–562. <https://doi.org/10.1127/1432-8364/2013/0200>.
- Bendig, Juliane, Andreas Bolten, Simon Bennertz, Janis Broscheit, Silas Eichfuss, and Georg Bareth. 2014. Estimating biomass of barley using crop surface models (CSMs) derived from UAV-based RGB imaging. *Remote Sensing* 6: 10395–10412. <https://doi.org/10.3390/rs61110395>.
- Bendig, Juliane, Kang Yu, Helge Aasen, Andreas Bolten, Simon Bennertz, Janis Broscheit, Martin L. Gnyp, and Georg Bareth. 2015. Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. *International Journal of Applied Earth Observation and Geoinformation* 39. Elsevier B.V.: 79–87. <https://doi.org/10.1016/j.jag.2015.02.012>.
- Berry, P. M. 1998. Predicting lodging in winter wheat. Ph.D. Thesis, University of Nottingham, Nottingham, United Kingdom.
- Berry, P. M., and J. Spink. 2012. Predicting yield losses caused by lodging in wheat. *Field Crops Research* 137. Elsevier B.V.: 19–26. <https://doi.org/10.1016/j.fcr.2012.07.019>.
- Berry, P. M., M. Sterling, J. H. Spink, C. J. Baker, R. Sylvester-Bradley, S. J. Mooney, A. R. Tams, and A. R. Ennos. 2004. Understanding and Reducing Lodging in Cereals. *Advances in Agronomy*. [https://doi.org/10.1016/S0065-2113\(04\)84005-7](https://doi.org/10.1016/S0065-2113(04)84005-7).
- Berteška, Tautvydas, and Birutė Ruzgienė. 2013. Photogrammetric mapping based on uav imagery. *Geodesy and Cartography*. <https://doi.org/10.3846/20296991.2013.859781>.
- Bláha, M., H. Eisenbeiss, D. Grimm, and P. Limpach. 2012. DIRECT GEOREFERENCING OF UAVS. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. <https://doi.org/10.5194/isprsarchives-XXXVIII-1-C22-131-2011>.
- Bondi, Elizabeth, Carl Salvaggio, Matthew Montanaro, and Aaron D. Gerace. 2016. Calibration of UAS imagery inside and outside of shadows for improved vegetation index computation. In *Autonomous Air and Ground Sensing Systems for Agricultural*

- Optimization and Phenotyping*. <https://doi.org/10.1117/12.2227214>.
- Briggs, K. G. 1990. STUDIES OF RECOVERY FROM ARTIFICIALLY INDUCED LODGING IN SEVERAL SIX-ROW BARLEY CULTIVARS. *Canadian Journal of Plant Science* 181: 173–181. <https://doi.org/10.4141/cjps90-019>.
- Burdziakowski, Pawel. 2018. Uav in todays photogrammetry – application areas and challenges. In *International Multidisciplinary Scientific GeoConference Surveying Geology and Mining Ecology Management, SGEM*. <https://doi.org/10.5593/sgem2018/2.3/S10.031>.
- Burkart, Andreas, Helge Aasen, Luis Alonso, Gunter Menz, Georg Bareth, and Uwe Rascher. 2015. Angular dependency of hyperspectral measurements over wheat characterized by a novel UAV based goniometer. *Remote Sensing* 7: 725–746. <https://doi.org/10.3390/rs70100725>.
- Busmeyer, Lucas, Daniel Mentrup, Kim Möller, Erik Wunder, Katharina Alheit, Volker Hahn, Hans Peter Maurer, et al. 2013. Breedvision - A multi-sensor platform for non-destructive field-based phenotyping in plant breeding. *Sensors (Switzerland)*. <https://doi.org/10.3390/s130302830>.
- Candiago, Sebastian, Fabio Remondino, Michaela De Giglio, Marco Dubbini, and Mario Gattelli. 2015. Evaluating Multispectral Images and Vegetation Indices for Precision Farming Applications from UAV Images. *Remote Sensing* 7: 4026–4047. <https://doi.org/10.3390/rs70404026>.
- Cao, Sen, Brad Danielson, Shari Clare, Shantel Koenig, Carlos Campos-Vargas, and Arturo Sanchez-Azofeifa. 2019. Radiometric calibration assessments for UAS-borne multispectral cameras: Laboratory and field protocols. *ISPRS Journal of Photogrammetry and Remote Sensing*. <https://doi.org/10.1016/j.isprsjprs.2019.01.016>.
- Casler, M. D., and K. P. Vogel. 1999. Accomplishments and impact from breeding for increased forage nutritional value. In *Crop Science*. <https://doi.org/10.2135/cropsci1999.0011183X003900010003x>.
- Chai, T., and R. R. Draxler. 2014. Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*. <https://doi.org/10.5194/gmd-7-1247-2014>.
- Challinor, A. J., J. Watson, D. B. Lobell, S. M. Howden, D. R. Smith, and N. Chhetri. 2014. A meta-analysis of crop yield under climate change and adaptation. *Nature Climate Change*. <https://doi.org/10.1038/nclimate2153>.
- Chapman, Scott, Torsten Merz, Amy Chan, Paul Jackway, Stefan Hrabar, M. Dreccer, Edward Holland, Bangyou Zheng, T. Ling, and Jose Jimenez-Berni. 2014. Pheno-Copter: A Low-Altitude, Autonomous Remote-Sensing Robotic Helicopter for High-Throughput Field-Based Phenotyping. *Agronomy* 4: 279–301. <https://doi.org/10.3390/agronomy4020279>.
- Chivasa, Walter, Onesimo Mutanga, and Chandrashekhar Biradar. 2020. UAV-based multispectral phenotyping for disease resistance to accelerate crop improvement under changing climate conditions. *Remote Sensing*. <https://doi.org/10.3390/RS12152445>.
- Choudhary, Piyush, AnandK. Chaudhari, Dr.A.N. Cheeran, and Sharada Godara. 2012. Color Transform Based Approach for Disease Spot Detection on Plant Leaf. *International Journal of Computer Science and Telecommunications*.
- Chu, Tianxing, Michael J. Starek, Michael J. Brewer, Seth C. Murray, and Luke S. Pruter. 2017. Assessing lodging severity over an experimental maize (*Zea mays* L.) field using UAS images. *Remote Sensing* 9: 1–24. <https://doi.org/10.3390/rs9090923>.
- Cohen, J. 1988. Statistical Power Analysis for the Behavioural Science (2<sup>nd</sup> ed). *Lawrence Erlbaum Associates*.
- Cointault, F., D. Guerin, J-P. Guillemin, and B. Chopinet. 2008. In-field triticum aestivum



- ear counting using colour texture image analysis. *New Zealand Journal of Crop and Horticultural Science* 36: 117–130. <https://doi.org/10.1080/01140670809510227>.
- Colomina, I., and P. Molina. 2014. Unmanned aerial systems for photogrammetry and remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*. <https://doi.org/10.1016/j.isprsjprs.2014.02.013>.
- Comba, L., A. Biglia, D. Ricauda Aimonino, C. Tortia, E. Mania, S. Guidoni, and P. Gay. 2020. Leaf Area Index evaluation in vineyards using 3D point clouds from UAV imagery. *Precision Agriculture*. <https://doi.org/10.1007/s11119-019-09699-x>.
- Core Team, R. 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Coumou, Dim, Alexander Robinson, and Stefan Rahmstorf. 2013. Global increase in record-breaking monthly-mean temperatures. *Climatic Change*. <https://doi.org/10.1007/s10584-012-0668-1>.
- Dahiya, Seema, Sandeep Kumar, and Charul Chaudhary. 2018. Lodging : Significance and preventive measures for increasing crop production. *International Journal of Chemical Studies* 6: 700–705.
- Dandois, Jonathan P., and Erle C. Ellis. 2013. High spatial resolution three-dimensional mapping of vegetation spectral dynamics using computer vision. *Remote Sensing of Environment*. <https://doi.org/10.1016/j.rse.2013.04.005>.
- Danielgm. 2016. CloudCompare: 3D point cloud and mesh processing software, Open Source Project. <http://www.danielgm.net/cc/>.
- De Souza, Carlos Henrique Wachholz, Rubens Augusto Camargo Lamparelli, Jansle Vieira Rocha, and Paulo Sergio Graziano Magalhães. 2017. Height estimation of sugarcane using an unmanned aerial system (UAS) based on structure from motion (SfM) point clouds. *International Journal of Remote Sensing*. <https://doi.org/10.1080/01431161.2017.1285082>.
- Deng, Jia, Wei Dong, R. Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. ImageNet: A large-scale hierarchical image database. *IEEE Conference on Computer Vision and Pattern Recognition*. <https://doi.org/10.1109/cvprw.2009.5206848>.
- Dorigo, W. A., R. Zurita-Milla, A. J.W. de Wit, J. Brazile, R. Singh, and M. E. Schaepman. 2007. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. *International Journal of Applied Earth Observation and Geoinformation*. <https://doi.org/10.1016/j.jag.2006.05.003>.
- Dos Santos, Gilberto Antonio Marcon, Zachary Barnes, Eric Lo, Bryan Ritoper, Lauren Nishizaki, Xavier Tejada, Alex Ke, et al. 2015. Small unmanned aerial vehicle system for wildlife radio collar tracking. In *Proceedings - 11th IEEE International Conference on Mobile Ad Hoc and Sensor Systems, MASS 2014*. <https://doi.org/10.1109/MASS.2014.48>.
- Doughty, Cheryl L., and Kyle C. Cavanaugh. 2019. Mapping coastal wetland biomass from high resolution unmanned aerial vehicle (UAV) imagery. *Remote Sensing*. <https://doi.org/10.3390/rs11050540>.
- Dronedeploy. 2017. DroneDeploy 's 2017 Drone Buyer 's Guide: 1–36. Retrived from <https://ncsu-geoforall-lab.github.io/uav-lidar-analytics-course/lectures/pdfs/DroneDeploy-Drone-Buyers-Guide-V1.3.pdf>, accessed January 02, 2022.
- Erisman, Jan Willem, Mark A. Sutton, James Galloway, Zbigniew Klimont, and Wilfried Winiwarter. 2008. How a century of ammonia synthesis changed the world. *Nature Geoscience*. <https://doi.org/10.1038/ngeo325>.
- Evenson, R. E., and D. Gollin. 2003. Assessing the impact of the Green Revolution, 1960 to 2000. *Science*. <https://doi.org/10.1126/science.1078710>.
- FAO, IFAD, UNICEF, WFP, and WHO. 2020. The State of Food Security and Nutrition in

- the World 2020. Transforming food systems for affordable healthy diets. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
- Fernandez-Gallego, Jose A., Ma Luisa Buchailot, Nieves Aparicio Gutiérrez, María Teresa Nieto-Taladriz, José Luis Araus, and Shawn C. Kefauver. 2019. Automatic wheat ear counting using thermal imagery. *Remote Sensing*. <https://doi.org/10.3390/rs11070751>.
- Fernandez-Gallego, Jose A., Shawn C. Kefauver, Nieves Aparicio Gutiérrez, María Teresa Nieto-Taladriz, and José Luis Araus. 2018. Wheat ear counting in-field conditions: High throughput and low-cost approach using RGB images. *Plant Methods*. <https://doi.org/10.1186/s13007-018-0289-4>.
- Fernandez-Gallego, Jose A., Peter Lootens, Irene Borra-Serrano, Veerle Derycke, Geert Haesaert, Isabel Roldán-Ruiz, Jose L. Araus, and Shawn C. Kefauver. 2020. Automatic wheat ear counting using machine learning based on RGB UAV imagery. *Plant Journal*. <https://doi.org/10.1111/tjp.14799>.
- Ferrante, Ariel, Julia Cartelle, Roxana Savin, and Gustavo A. Slafer. 2017. Yield determination, interplay between major components and yield stability in a traditional and a contemporary wheat across a wide range of environments. *Field Crops Research*. <https://doi.org/10.1016/j.fcr.2016.12.028>.
- Ferreira, dos Santos Alessandro, Daniel Matte Freitas, Gercina Gonçalves da Silva, Hemerson Pistori, and Marcelo Theophilo Folhes. 2017. Weed detection in soybean crops using ConvNets. *Computers and Electronics in Agriculture*. <https://doi.org/10.1016/j.compag.2017.10.027>.
- Fischer, R. A., and M. Stapper. 1987. Lodging effects on high-yielding crops of irrigated semidwarf wheat. *Field Crops Research* 17: 245–258. [https://doi.org/10.1016/0378-4290\(87\)90038-4](https://doi.org/10.1016/0378-4290(87)90038-4).
- Furbank, Robert T., and Mark Tester. 2011. Phenomics - technologies to relieve the phenotyping bottleneck. *Trends in Plant Science*. <https://doi.org/10.1016/j.tplants.2011.09.005>.
- Gao, Ge, Mark A. Tester, and Magdalena M. Julkowska. 2020. The Use of High-Throughput Phenotyping for Assessment of Heat Stress-Induced Changes in Arabidopsis. *Plant Phenomics*. <https://doi.org/10.34133/2020/3723916>.
- García-Martínez, Héctor, Héctor Flores-Magdaleno, Abdul Khalil-Gardezi, Roberto Ascencio-Hernández, Leonardo Tijerina-Chávez, Mario A. Vázquez-Peña, and Oscar R. Mancilla-Villa. 2020. Digital count of corn plants using images taken by unmanned aerial vehicles and cross correlation of templates. *Agronomy*. <https://doi.org/10.3390/agronomy10040469>.
- García Del Moral, L. F., Y. Rharrabti, D. Villegas, and C. Royo. 2003. Evaluation of grain yield and its components in durum wheat under Mediterranean conditions: An ontogenic approach. *Agronomy Journal*. <https://doi.org/10.2134/agronj2003.2660>.
- Gebbers, Robin, and Viacheslav I. Adamchuk. 2010. Precision agriculture and food security. *Science*. <https://doi.org/10.1126/science.1183899>.
- Gómez-Candón, D., A. I. De Castro, and F. López-Granados. 2014. Assessing the accuracy of mosaics from unmanned aerial vehicle (UAV) imagery for precision agriculture purposes in wheat. *Precision Agriculture* 15: 44–56. <https://doi.org/10.1007/s11119-013-9335-4>.
- Gong, Chizhang, Henning Buddenbaum, Rebecca Retzlaff, and Thomas Udelhoven. 2019. An empirical assessment of angular dependency for RedEdge-M in sloped terrain viticulture. *Remote Sensing*. <https://doi.org/10.3390/rs11212561>.
- Gonzalez, Luis F., Glen A. Montes, Eduard Puig, Sandra Johnson, Kerrie Mengersen, and Kevin J. Gaston. 2016. Unmanned aerial vehicles (UAVs) and artificial intelligence revolutionizing wildlife monitoring and conservation. *Sensors (Switzerland)*.

- <https://doi.org/10.3390/s16010097>.
- Gülch, E. 2012. PHOTOGAMMETRIC MEASUREMENTS IN FIXED WING UAV IMAGERY. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. <https://doi.org/10.5194/isprsarchives-xxxix-b1-381-2012>.
- Hakala, Teemu, Lauri Markelin, Eija Honkavaara, Barry Scott, Theo Theocharous, Olli Nevalainen, Roope Näsi, et al. 2018. Direct reflectance measurements from drones: Sensor absolute radiometric calibration and system tests for forest reflectance characterization. *Sensors (Switzerland)*. <https://doi.org/10.3390/s18051417>.
- Hansen, P. M., and J. K. Schjoerring. 2003. Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. *Remote Sensing of Environment*. [https://doi.org/10.1016/S0034-4257\(03\)00131-7](https://doi.org/10.1016/S0034-4257(03)00131-7).
- Harvey, Mark, and Sarah Pilgrim. 2011. The new competition for land: Food, energy, and climate change. *Food Policy*. <https://doi.org/10.1016/j.foodpol.2010.11.009>.
- Hasan, Md Mehedi, Joshua P. Chopin, Hamid Laga, and Stanley J. Miklavcic. 2018. Detection and analysis of wheat spikes using Convolutional Neural Networks. *Plant Methods*. <https://doi.org/10.1186/s13007-018-0366-8>.
- Hassan, Muhammad Adeel, Mengjiao Yang, Awais Rasheed, Guijun Yang, Matthew Reynolds, Xianchun Xia, Yonggui Xiao, and Zhonghu He. 2019. A rapid monitoring of NDVI across the wheat growth cycle for grain yield prediction using a multi-spectral UAV platform. *Plant Science*. <https://doi.org/10.1016/j.plantsci.2018.10.022>.
- Henchion, Maeve, Mary McCarthy, Virginia C. Resconi, and Declan Troy. 2014. Meat consumption: Trends and quality matters. *Meat Science*. <https://doi.org/10.1016/j.meatsci.2014.06.007>.
- Hirt, Christian. 2017. Digital Terrain Models. *Encyclopedia of Geodesy*: 0–6. <https://doi.org/10.1007/978-3-319-02370-0>.
- Holman, Fenner H., Andrew B. Riche, Adam Michalski, March Castle, Martin J. Wooster, and Malcolm J. Hawkesford. 2016. High throughput field phenotyping of wheat plant height and growth rate in field plot trials using UAV based remote sensing. *Remote Sensing* 8. <https://doi.org/10.3390/rs8121031>.
- Homolová, Lucie, Zbyněk Malenovský, Jan G.P.W. Clevers, Glenda García-Santos, and Michael E. Schaepman. 2013. Review of optical-based remote sensing for plant trait mapping. *Ecological Complexity*. <https://doi.org/10.1016/j.ecocom.2013.06.003>.
- Horstrand, P., R. Guerra, M. Diaz, A. Morales, A. Jimenez, S. Lopez, and J. F. Lopez. 2019. A spectral imaging system for precision agriculture: From its inception till a pre-commercial prototype. In *2019 34th Conference on Design of Circuits and Integrated Systems, DCIS 2019*. <https://doi.org/10.1109/DCIS201949030.2019.8959891>.
- Houborg, Rasmus, and Matthew F. McCabe. 2018. A hybrid training approach for leaf area index estimation via Cubist and random forests machine-learning. *ISPRS Journal of Photogrammetry and Remote Sensing*. <https://doi.org/10.1016/j.isprsjprs.2017.10.004>.
- Houghton, J.T., Y. Ding, D.J. Griggs, M. Noguer, P.J. van der Linden, X. Dai, K. Maskell, and C.A. Johnson. 2001. CLIMATECHANGE 2001: THE SCIENTIFIC BASIS. *Cambridge University Press*.
- Hunt, E. Raymond, Paul C. Doraiswamy, James E. McMurtrey, Craig S.T. Daughtry, Eileen M. Perry, and Bakhyt Akhmedov. 2012. A visible band index for remote sensing leaf chlorophyll content at the Canopy scale. *International Journal of Applied Earth Observation and Geoinformation*. <https://doi.org/10.1016/j.jag.2012.07.020>.
- Ittersum, M. K., S.M. Howden, S. Asseng. 2003. Sensitivity of productivity and deep drainage of wheat cropping systems in a Mediterranean environment to changes in

- CO<sub>2</sub>, temperature and precipitation. *Agriculture, Ecosystems and Environment*.  
[https://doi.org/10.1016/S0167-8809\(03\)00114-2](https://doi.org/10.1016/S0167-8809(03)00114-2).
- Jackson, R. D., J. L. Hatfield, R. J. Reginato, S. B. Idso, and P. J. Pinter. 1983. Estimation of daily evapotranspiration from one time-of-day measurements. *Agricultural Water Management*. [https://doi.org/10.1016/0378-3774\(83\)90095-1](https://doi.org/10.1016/0378-3774(83)90095-1).
- Jaud, Marion, Sophie Passot, Pascal Allemand, Nicolas Le Dantec, Philippe Grandjean, and Christophe Delacourt. 2019. Suggestions to limit geometric distortions in the reconstruction of linear coastal landforms by sfm photogrammetry with photoscan® and micmac® for uav surveys with restricted GCPs pattern. *Drones*.  
<https://doi.org/10.3390/drones3010002>.
- Jin, Xiuliang, Zhenhai Li, Guijun Yang, Hao Yang, Haikuan Feng, Xingang Xu, Jihua Wang, Xinchuan Li, and Juhua Luo. 2017a. Winter wheat yield estimation based on multi-source medium resolution optical and radar imaging data and the AquaCrop model using the particle swarm optimization algorithm. *ISPRS Journal of Photogrammetry and Remote Sensing*. <https://doi.org/10.1016/j.isprsjprs.2017.02.001>.
- Jin, Xiuliang, Shouyang Liu, Frédéric Baret, Matthieu Hemerlé, and Alexis Comar. 2017b. Estimates of plant density of wheat crops at emergence from very low altitude UAV imagery. *Remote Sensing of Environment* 198: 105–114.  
<https://doi.org/10.1016/j.rse.2017.06.007>.
- Joalland, Samuel, Claudio Screpanti, Hubert Vincent Varella, Marie Reuther, Mareike Schwind, Christian Lang, Achim Walter, and Frank Liebisch. 2018. Aerial and ground based sensing of tolerance to beet cyst nematode in sugar beet. *Remote Sensing*. <https://doi.org/10.3390/rs10050787>.
- Joseph, K. D. S. M., M. M. Alley, D. E. Brann, and W. D. Gravelle. 1985. Row Spacing and Seeding Rate Effects on Yield and Yield Components of Soft Red Winter Wheat. *Agronomy Journal*. <https://doi.org/10.2134/agronj1985.00021962007700020009x>.
- Kamilaris, A., and F. X. Prenafeta-Boldú. 2018. A review of the use of convolutional neural networks in agriculture. *Journal of Agricultural Science*.  
<https://doi.org/10.1017/S0021859618000436>.
- Kersten, Thomas P. 2016. Auf Knopfdruck – Auswertung und Resultate UAV-gestützter Bildflugdaten. *UAV 2016 - Vermessung mit unbemannten Flugsystemen, Schriftenreihe des DVW* 82: 135–158.
- Kirby, E. J.M. 1993. Effect of sowing depth on seedling emergence, growth and development in barley and wheat. *Field Crops Research*.  
[https://doi.org/10.1016/0378-4290\(93\)90143-B](https://doi.org/10.1016/0378-4290(93)90143-B).
- Kittler, J., and J. Illingworth. 1985. On Threshold Selection Using Clustering Criteria. *IEEE Transactions on Systems, Man and Cybernetics*.  
<https://doi.org/10.1109/TSMC.1985.6313443>.
- Koh, Joshua C.O., Matthew Hayden, Hans Daetwyler, and Surya Kant. 2019. Estimation of crop plant density at early mixed growth stages using UAV imagery. *Plant Methods*.  
<https://doi.org/10.1186/s13007-019-0449-1>.
- Kraus, K., C. Briese, M. Attwenger, N. Preifer. 2004. Quality measures for digital terrain models. *International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences*.
- Kulkarni, Anand H, and Ashwin Patil R K. 2012. Applying image processing technique to detect plant diseases. *International Journal of Modern Engineering Research (IJMER)*.
- Kunz, A., and Völkerling G. 2015. Feldversuche Campus Klein-Altendorf. Retrived from <https://www.cka.uni-bonn.de/versuchsfeldfuehrer-2015>, accessed January 02, 2022.
- Kurniawati, Nunik Noviana, Siti Norul Huda Sheikh Abdullah, Salwani Abdullah, and Saad Abdullah. 2009. Investigation on image processing techniques for diagnosing

- paddy diseases. In *SoCPaR 2009 - Soft Computing and Pattern Recognition*.  
<https://doi.org/10.1109/SoCPaR.2009.62>.
- Lawles, Kyle, William Raun, Kefyalew Desta, and Kyle Freeman. 2012. Effect of delayed emergence on corn grain yields. *Journal of Plant Nutrition*.  
<https://doi.org/10.1080/01904167.2012.639926>.
- Lerman, Jerrold. 1996. Study design in clinical research: Sample size estimation and power analysis. *Canadian Journal of Anaesthesia*. <https://doi.org/10.1007/BF03011261>.
- Li, Bo, Xiangming Xu, Jiwan Han, Li Zhang, Chunsong Bian, Liping Jin, and Jiangang Liu. 2019. The estimation of crop emergence in potatoes by UAV RGB imagery. *Plant Methods* 15: 1–13. <https://doi.org/10.1186/s13007-019-0399-7>.
- Li, Zhilin. 1988. ON THE MEASURE OF DIGITAL TERRAIN MODEL ACCURACY. *The Photogrammetric Record*. <https://doi.org/10.1111/j.1477-9730.1988.tb00636.x>.
- Liaghat, S., and S. K. Balasundram. 2010. A review: The role of remote sensing in precision agriculture. *American Journal of Agricultural and Biological Science*.  
<https://doi.org/10.3844/ajabssp.2010.50.55>.
- Liakos, Konstantinos G., Patrizia Busato, Dimitrios Moshou, Simon Pearson, and Dionysis Bochtis. 2018. Machine learning in agriculture: A review. *Sensors (Switzerland)*.  
<https://doi.org/10.3390/s18082674>.
- Lindstrom, M. J., R. I. Papendick, and F. E. Koehler. 1976. A Model to Predict Winter Wheat Emergence as Affected by Soil Temperature, Water Potential, and Depth of Planting. *Agronomy Journal*.  
<https://doi.org/10.2134/agronj1976.00021962006800010038x>.
- Liu, Shouyang, Fred Baret, Denis Allard, Xiuliang Jin, Bruno Andrieu, Philippe Burger, Matthieu Hemmerlé, and Alexis Comar. 2017a. A method to estimate plant density and plant spacing heterogeneity: Application to wheat crops. *Plant Methods*: 1–11.  
<https://doi.org/10.1186/s13007-017-0187-1>.
- Liu, Shouyang, Fred Baret, Bruno Andrieu, Philippe Burger, and Matthieu Hemmerlé. 2017b. Estimation of wheat plant density at early stages using high resolution imagery. *Frontiers in Plant Science* 8: 1–10. <https://doi.org/10.3389/fpls.2017.00739>.
- Liu, Tao, Rui Li, Xiaochun Zhong, Min Jiang, Xiuliang Jin, Ping Zhou, Shengping Liu, Chengming Sun, and Wenshan Guo. 2018. Estimates of rice lodging using indices derived from UAV visible and thermal infrared images. *Agricultural and Forest Meteorology* 252. Elsevier: 144–154. <https://doi.org/10.1016/j.agrformet.2018.01.021>.
- Liu, Tao, Chengming Sun, Lijian Wang, Xiaochun Zhong, Xinkai Zhu, and Wenshan Guo. 2014. In-field wheatear counting based on image processing technology. *Nongye Jixie Xuebao/Transactions of the Chinese Society for Agricultural Machinery*.  
<https://doi.org/10.6041/j.issn.1000-1298.2014.02.047>.
- López-Granados, Francisca, Jorge Torres-Sánchez, Angélica Serrano-Pérez, Ana I. de Castro, Fco Javier Mesas-Carrascosa, and José Manuel Peña. 2016. Early season weed mapping in sunflower using UAV technology: variability of herbicide treatment maps against weed thresholds. *Precision Agriculture*. <https://doi.org/10.1007/s11119-015-9415-8>.
- Lottes, Philipp, Raghav Khanna, Johannes Pfeifer, Roland Siegwart, and Cyrill Stachniss. 2017. UAV-based crop and weed classification for smart farming. In *Proceedings - IEEE International Conference on Robotics and Automation*.  
<https://doi.org/10.1109/ICRA.2017.7989347>.
- Lowe, David G. 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60: 91–110.  
<https://doi.org/10.1023/B:VISI.0000029664.99615.94>.
- Lu, Ning, Jie Zhou, Zixu Han, Dong Li, Qiang Cao, Xia Yao, Yongchao Tian, Yan Zhu, Weixing Cao, and Tao Cheng. 2019. Improved estimation of aboveground biomass in

- wheat from RGB imagery and point cloud data acquired with a low-cost unmanned aerial vehicle system. *Plant Methods*. <https://doi.org/10.1186/s13007-019-0402-3>.
- Macedo-Cruz, Antonia, Gonzalo Pajares, Matilde Santos, and Isidro Villegas-Romero. 2011. Digital image sensor-based assessment of the status of oat (*Avena sativa* L.) crops after frost damage. *Sensors*. <https://doi.org/10.3390/s110606015>.
- Madec, Simon, Fred Baret, Benoît De Solan, Samuel Thomas, Dan Dutartre, Stéphane Jezequel, Matthieu Hemmerlé, Gallian Colombeau, and Alexis Comar. 2017. High-throughput phenotyping of plant height: Comparing unmanned aerial vehicles and ground lidar estimates. *Frontiers in Plant Science*. <https://doi.org/10.3389/fpls.2017.02002>.
- Madec, Simon, Xiuliang Jin, Hao Lu, Benoit De Solan, Shouyang Liu, Florent Duyme, Emmanuelle Heritier, and Frédéric Baret. 2019. Ear density estimation from high resolution RGB imagery using deep learning technique. *Agricultural and Forest Meteorology*. <https://doi.org/10.1016/j.agrformet.2018.10.013>.
- Mamaghani, Baabak, Ryan Connal, Ryan Hartzell, Kevin Kha, Geoffrey Sasaki, Evan Marcellus, Jackson Knappen, Timothy Bauch, Nina Raqueno, and Carl Salvaggio. 2018. An initial exploration of vicarious and in-scene calibration techniques for small unmanned aircraft systems. *Commercial & Scientific Sensing and Imaging*. <https://doi.org/10.1117/12.2302744>.
- Marcial-Pablo, Mariana de Jesús, Alberto Gonzalez-Sanchez, Sergio Iván Jimenez-Jimenez, Ronald Ernesto Ontiveros-Capurata, and Waldo Ojeda-Bustamante. 2019. Estimation of vegetation fraction using RGB and multispectral images from UAV. *International Journal of Remote Sensing*. <https://doi.org/10.1080/01431161.2018.1528017>.
- Meier, Uwe. 2001. Growth stages of mono- and dicotyledonous plants. *Federal Biological Research Centre for Agriculture and Forestry*. <https://doi.org/10.5073/bbch0515>.
- Messina, Gaetano, and Giuseppe Modica. 2020. Applications of UAV thermal imagery in precision agriculture: State of the art and future research outlook. *Remote Sensing*. <https://doi.org/10.3390/RS12091491>.
- Meyer, George E., and João Camargo Neto. 2008. Verification of color vegetation indices for automated crop imaging applications. *Computers and Electronics in Agriculture*. <https://doi.org/10.1016/j.compag.2008.03.009>.
- MicaSense. 2017. Light Sensors: the Basics (DLS and Sunshine Sensor). Retrived from <https://support.micasense.com/hc/en-us/articles/115002782008-Light-Sensors-the-Basics-DLS-and-Sunshine-Sensor->, accessed January 02, 2022.
- Mohan, Midhun, Carlos Alberto Silva, Carine Klauberg, Prahlad Jat, Glenn Catts, Adrián Cardil, Andrew Thomas Hudak, and Mahendra Dia. 2017. Individual tree detection from unmanned aerial vehicle (UAV) derived canopy height model in an open canopy mixed conifer forest. *Forests*. <https://doi.org/10.3390/f8090340>.
- Morota, Gota, Ricardo V. Ventura, Fabyano F. Silva, Masanori Koyama, and Samodha C. Fernando. 2018. Big data analytics and precision animal agriculture symposium: Machine learning and data mining advance predictive big data analysis in precision animal agriculture. *Journal of Animal Science*. <https://doi.org/10.1093/jas/sky014>.
- Mulla, David J. 2013. Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems Engineering* 114. IAgRE: 358–371. <https://doi.org/10.1016/j.biosystemseng.2012.08.009>.
- Murakami, Toshifumi, Mamiko Yui, and Koichi Amaha. 2012. Canopy height measurement by photogrammetric analysis of aerial images: Application to buckwheat (*Fagopyrum esculentum* Moench) lodging evaluation. *Computers and Electronics in Agriculture* 89. Elsevier B.V.: 70–75. <https://doi.org/10.1016/j.compag.2012.08.003>.

- Murungu, F. S. 2011. Effects of seed priming and water potential on seed germination and emergence of wheat (*Triticum aestivum* L.) varieties in laboratory assays and in the field. *African Journal of Biotechnology*. <https://doi.org/10.5897/AJB10.1630>.
- Navabi, A, M Iqbal, K Strenzke, and D Spaner. 2006. The relationship between lodging and plant height in a diverse wheat population. *Canadian Journal of Plant Science* 86: 723–726. <https://doi.org/10.4141/P05-144>.
- Nebiker, S., N. Lack, M. Abächerli, and S. Läderach. 2016. Light-weight multispectral uav sensors and their capabilities for predicting grain yield and detecting plant diseases. In *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*. <https://doi.org/10.5194/isprsarchives-XLI-B1-963-2016>.
- Neto, Joao Camargo. 2010. a combined statistical-soft computing approach for classification and mapping weed species in minimum-tillage systems. *Cornell Hotel and Restaurant Administration Quarterly*. <https://doi.org/10.1177/001088048102200214>.
- Ollero, A., J.R. Martínez-de-Dios, and L. Merino. 2006. Unmanned aerial vehicles as tools for forest-fire fighting. *Forest Ecology and Management*. <https://doi.org/10.1016/j.foreco.2006.08.292>.
- Olsen, Jannie, Lars Kristensen, and Jacob Weiner. 2006. Influence of sowing density and spatial pattern of spring wheat (*Triticum aestivum*) on the suppression of different weed species. *Weed Biology and Management*. <https://doi.org/10.1111/j.1445-6664.2006.00210.x>.
- Olsen, Jannie, and Jacob Weiner. 2007. The influence of *Triticum aestivum* density, sowing pattern and nitrogen fertilization on leaf area index and its spatial variation. *Basic and Applied Ecology*. <https://doi.org/10.1016/j.baae.2006.03.013>.
- Otsu, Nobuyuki. 1979. THRESHOLD SELECTION METHOD FROM GRAY-LEVEL HISTOGRAMS. *IEEE Trans Syst Man Cybern*. <https://doi.org/10.1109/tsmc.1979.4310076>.
- Özyeşil, Onur, Vladislav Voroninski, Ronen Basri, and Amit Singer. 2017. A survey of structure from motion. *Acta Numerica*. <https://doi.org/10.1017/S096249291700006X>.
- Pask, Alistair, Julian Pietragalla, and Debra Mullan. 2012. Physiological Breeding II: A Field Guide to Wheat Phenotyping. *International Maize and Wheat Improvement Center (CIMMYT)*.
- Peng, Dianliang, Xiaoguang Chen, Yanping Yin, Kunli Lu, Weibing Yang, Yuhai Tang, and Zhenlin Wang. 2014. Lodging resistance of winter wheat (*Triticum aestivum* L.): Lignin accumulation and its related enzymes activities due to the application of paclobutrazol or gibberellin acid. *Field Crops Research*. <https://doi.org/10.1016/j.fcr.2013.11.015>.
- Pingali, Prabhu L. 2012. Green revolution: Impacts, limits, and the path ahead. *Proceedings of the National Academy of Sciences of the United States of America*. <https://doi.org/10.1073/pnas.0912953109>.
- Pinthus, Moshe J. 1974. Lodging in Wheat, Barley, and Oats: The Phenomenon, its Causes, and Preventive Measures. *Advances in Agronomy*. [https://doi.org/10.1016/S0065-2113\(08\)60782-8](https://doi.org/10.1016/S0065-2113(08)60782-8).
- Podobnikar, T. 2009. Methods for visual quality assessment of a digital terrain model. *S.a.P.I.En.S* 1: 1–24.
- Postma, Johannes A, Vera L Hecht, Kouki Hikosaka, Eric A Nord, and Pons Hendrik Poorter. 2020. Dividing the pie : A quantitative review on plant density responses. *Plant, Cell & Environment*: 0–2. <https://doi.org/10.1111/pce.13968>.
- Prasad, P. V. V., S. A. Staggenborg, and Z. Ristic. 2015. Impacts of Drought and/or Heat Stress on Physiological, Developmental, Growth, and Yield Processes of Crop Plants.



- Advances in Agricultural Systems Modeling*.  
<https://doi.org/10.2134/advagricsystmodell.c11>.
- Przybilla, H. J., M. Bäumker, T. Luhmann, H. Hastedt, and M. Eilers. 2020. Interaction between direct georeferencing, control point configuration and camera self-calibration for RTK-based UAV photogrammetry. In *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*. <https://doi.org/10.5194/isprs-archives-XLIII-B1-2020-485-2020>.
- Rajala, Ari, Markku Niskanen, Mika Isoaho, and Pirjo Peltonen-Sainio. 2011. Seed quality effects on seedling emergence, plant stand establishment and grain yield in two-row barley. *Agricultural and Food Science*.  
<https://doi.org/10.2137/145960611797471516>.
- Rajapaksa, Sajith, Mark Eramian, Hema Duddu, Menglu Wang, Steve Shirtliffe, Seungbum Ryu, Anique Josuttes, et al. 2018. Classification of Crop Lodging with Gray Level Co-occurrence Matrix. *Proceedings - 2018 IEEE Winter Conference on Applications of Computer Vision, WACV 2018*: 251–258.  
<https://doi.org/10.1109/WACV.2018.00034>.
- Rathmann, Régis, Alexandre Szklo, and Roberto Schaeffer. 2010. Land use competition for production of food and liquid biofuels: An analysis of the arguments in the current debate. *Renewable Energy*. <https://doi.org/10.1016/j.renene.2009.02.025>.
- Richardson, Andrew D., Julian P. Jenkins, Bobby H. Braswell, David Y. Hollinger, Scott V. Ollinger, and Marie Louise Smith. 2007. Use of digital webcam images to track spring green-up in a deciduous broadleaf forest. *Oecologia*.  
<https://doi.org/10.1007/s00442-006-0657-z>.
- Riehle, Daniel, David Reiser, and Hans W. Griepentrog. 2020. Robust index-based semantic plant/background segmentation for RGB- images. *Computers and Electronics in Agriculture*. <https://doi.org/10.1016/j.compag.2019.105201>.
- Sadeghi-Tehran, Pouria, Nicolas Virlet, Eva M. Ampe, Piet Reyns, and Malcolm J. Hawkesford. 2019. DeepCount: In-Field Automatic Quantification of Wheat Spikes Using Simple Linear Iterative Clustering and Deep Convolutional Neural Networks. *Frontiers in Plant Science*. <https://doi.org/10.3389/fpls.2019.01176>.
- Samarah, Nezar H. 2005. Effects of drought stress on growth and yield of barley. *Agronomie*. <https://doi.org/10.1051/agro:2004064>.
- Sankaran, Sindhuja, Lav R. Khot, and Arron H. Carter. 2015. Field-based crop phenotyping: Multispectral aerial imaging for evaluation of winter wheat emergence and spring stand. *Computers and Electronics in Agriculture* 118. Elsevier B.V.: 372–379. <https://doi.org/10.1016/j.compag.2015.09.001>.
- Sankaran, Sindhuja, Juan José Quirós, N. Richard Knowles, and Lisa O. Knowles. 2017. High-Resolution Aerial Imaging Based Estimation of Crop Emergence in Potatoes. *American Journal of Potato Research*: 658–663. <https://doi.org/10.1007/s12230-017-9604-2>.
- Scherer, Jürgen, Saeed Yahyanejad, Samira Hayat, Evşen Yanmaz, Vladimir Vukadinovic, Torsten Andre, Christian Bettstetter, Bernhard Rinner, Asif Khan, and Hermann Hellwagner. 2015. An autonomous multi-UAV system for search and rescue. In *DroNet 2015 - Proceedings of the 2015 Workshop on Micro Aerial Vehicle Networks, Systems, and Applications for Civilian Use*. <https://doi.org/10.1145/2750675.2750683>.
- Schiermeier, Quirin. 2019. Climate change made Europe’s mega-heatwave five times more likely. *Nature*. <https://doi.org/10.1038/d41586-019-02071-z>.
- Schneider, Caroline A., Wayne S. Rasband, and Kevin W. Eliceiri. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*.  
<https://doi.org/10.1038/nmeth.2089>.
- Shakhatreh, Hazim, Ahmad H. Sawalmeh, Ala Al-Fuqaha, Zuochoao Dou, Eyad Almaita,

- Issa Khalil, Noor Shamsiah Othman, Abdallah Khreishah, and Mohsen Guizani. 2019. Unmanned Aerial Vehicles (UAVs): A Survey on Civil Applications and Key Research Challenges. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2019.2909530>.
- Shanahan, J. F., J. S. Schepers, D. D. Francis, G. E. Varvel, W. W. Wilhelm, J. M. Tringe, M. R. Schlemmer, and D. J. Major. 2001. Use of remote-sensing imagery to estimate corn grain yield. *Agronomy Journal*. <https://doi.org/10.2134/agronj2001.933583x>.
- Shrestha, Dev Sagar, and Brian L. Steward. 2003. Automatic corn plant population measurement using machine vision. *Transactions of the American Society of Agricultural Engineers* 46: 559–565. <https://doi.org/10.13031/2013.12945>.
- Sieberth, T., R. Wackrow, and J. H. Chandler. 2014. Motion blur disturbs - the influence of motion-blurred images in photogrammetry. *Photogrammetric Record*. <https://doi.org/10.1111/phor.12082>.
- Simko, Ivan, and Hans Peter Piepho. 2011. Combining phenotypic data from ordinal rating scales in multiple plant experiments. *Trends in Plant Science*. <https://doi.org/10.1016/j.tplants.2011.02.001>.
- Slafer, Gustavo A., Roxana Savin, and Victor O. Sadras. 2014. Coarse and fine regulation of wheat yield components in response to genotype and environment. *Field Crops Research*. <https://doi.org/10.1016/j.fcr.2013.12.004>.
- Slafer, Gustavo A, D F Calderini, and D J Miralles. 1996. Yield components and compensation in wheat: opportunities for further increasing yield potential. *CIMMYT*.
- Smith, Geoffrey M., and Edward J. Milton. 1999. The use of the empirical line method to calibrate remotely sensed data to reflectance. *International Journal of Remote Sensing*. <https://doi.org/10.1080/014311699211994>.
- Soares, Gabriel, Leonardo Campos Inocencio, Maurício Roberto Veronez, Luiz Gonzaga Da Silveira, Fabiane Bordin, and Fernando Pinho Marson. 2018. Analysis of positional and geometric accuracy of objects in survey with unmanned aerial vehicle (UAV). In *International Geoscience and Remote Sensing Symposium (IGARSS)*. <https://doi.org/10.1109/IGARSS.2018.8517315>.
- Solomon, S., D., M. Qin, Z. Manning, M. Chen, K.B. Marquis, M. Tignor Averyt, Miller HL, et al. 2007. Summary for Policymakers. In: *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.
- Song, Yang, and Jinfei Wang. 2019. Winter wheat canopy height extraction from UAV-based point cloud data with a moving cuboid filter. *Remote Sensing*. <https://doi.org/10.3390/rs11101239>.
- Sonnentag, O., M. Detto, R. Vargas, Y. Ryu, B. R.K. Runkle, M. Kelly, and D. D. Baldocchi. 2011. Tracking the structural and functional development of a perennial pepperweed (*Lepidium latifolium* L.) infestation using a multi-year archive of webcam imagery and eddy covariance measurements. *Agricultural and Forest Meteorology*. <https://doi.org/10.1016/j.agrformet.2011.02.011>.
- Srbínovska, Mare, Cvetan Gavrovski, Vladimir Dimcev, Aleksandra Krkoleva, and Vesna Borozan. 2015. Environmental parameters monitoring in precision agriculture using wireless sensor networks. *Journal of Cleaner Production*. <https://doi.org/10.1016/j.jclepro.2014.04.036>.
- Stanton, Carly, Michael J. Starek, Norman Elliott, Michael Brewer, Murilo M. Maeda, and Tianxing Chu. 2017. Unmanned aircraft system-derived crop height and normalized difference vegetation index metrics for sorghum yield and aphid stress assessment. *Journal of Applied Remote Sensing*. <https://doi.org/10.1117/1.JRS.11.026035>.
- Steiner, J. J., D. F. Grabe, and M. Tulo. 1989. Single and Multiple Vigor Tests for Predicting Seedling Emergence of Wheat. *Crop Science*. <https://doi.org/10.2135/cropsci1989.0011183x002900030049x>.

- Stott, Eilidh, Richard D. Williams, and Trevor B. Hoey. 2020. Ground control point distribution for accurate kilometre-scale topographic mapping using an rtk-gnss unmanned aerial vehicle and sfm photogrammetry. *Drones*. <https://doi.org/10.3390/drones4030055>.
- Štroner, Martin, Rudolf Urban, Tomáš Reindl, Jan Seidl, and Josef Brouček. 2020. Evaluation of the georeferencing accuracy of a photogrammetric model using a quadrocopter with onboard GNSS RTK. *Sensors (Switzerland)*. <https://doi.org/10.3390/s20082318>.
- Su, Jinya, Cunjia Liu, Xiaoping Hu, Xiangming Xu, Lei Guo, and Wen Hua Chen. 2019. Spatio-temporal monitoring of wheat yellow rust using UAV multispectral imagery. *Computers and Electronics in Agriculture*. <https://doi.org/10.1016/j.compag.2019.105035>.
- Suomalainen, Juha, Teemu Hakala, Raquel Alves de Oliveira, Lauri Markelin, Niko Viljanen, Roope Näsi, and Eija Honkavaara. 2018. A novel tilt correction technique for irradiance sensors and spectrometers on-board unmanned aerial vehicles. *Remote Sensing*. <https://doi.org/10.3390/rs10122068>.
- Susko, Alexander Q., Fletcher Gilbertson, D. Jo Heuschele, Kevin Smith, and Peter Marchetto. 2018. An automatable, field camera track system for phenotyping crop lodging and crop movement. *HardwareX*. <https://doi.org/10.1016/j.ohx.2018.e00029>.
- Süsstrunk, Sabine, Robert Buckley, and Steve Swen. 1999. Standard RGB color spaces. In *Final Program and Proceedings - IS and T/SID Color Imaging Conference*.
- Taddia, Yuri, Paolo Russo, Stefano Lovo, and Alberto Pellegrinelli. 2020. Multispectral UAV monitoring of submerged seaweed in shallow water. *Applied Geomatics*. <https://doi.org/10.1007/s12518-019-00270-x>.
- Tait, Leigh, Jochen Bind, Hannah Charan-Dixon, Ian Hawes, John Pirker, and David Schiel. 2019. Unmanned aerial vehicles (UAVs) for monitoring macroalgal biodiversity: Comparison of RGB and multispectral imaging sensors for biodiversity assessments. *Remote Sensing*. <https://doi.org/10.3390/rs11192332>.
- Tang, Lie, and Lei F. Tian. 2008. Plant identification in mosaicked crop row images for automatic emerged corn plant spacing measurement. *Transactions of the ASABE* 51: 2181–2191.
- Tekin, Arif Behiç. 2010. Variable rate fertilizer application in Turkish wheat agriculture: Economic assessment. *African Journal of Agricultural Research*. <https://doi.org/10.5897/AJAR09.562>.
- Tellaeche, Alberto, Xavier P. BurgosArtizzu, Gonzalo Pajares, Angela Ribeiro, and César Fernández-Quintanilla. 2008. A new vision-based approach to differential spraying in precision agriculture. *Computers and Electronics in Agriculture*. <https://doi.org/10.1016/j.compag.2007.07.008>.
- Teppati Losè, L., F. Chiabrandò, and F. Giulio Tonolo. 2020. Are measured ground control points still required in UAV based large scale mapping? assessing the positional accuracy of an RTK multi-rotor platform. In *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*. <https://doi.org/10.5194/isprs-archives-XLIII-B1-2020-507-2020>.
- Tewes, Andreas. 2018. Investigating the Potential of UAV-Based Low-Cost Camera Imagery for Measuring Biophysical Variables in Maize. Ph.D. Thesis, Rheinischen Friedrich-Wilhelms-Universität Bonn, Bonn, Germany.
- Thomasson, John A., Cody Bagnall, Chao Sima, and Chenghai Yang. 2018. Quality assessment of radiometric calibration of UAV image mosaics. In: *Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping III* 10664: 1066404. <https://doi.org/10.1117/12.2305635>.
- Tian, Jinyan, Le Wang, Xiaojuan Li, Huili Gong, Chen Shi, Ruofei Zhong, and Xiaomeng

- Liu. 2017. Comparison of UAV and WorldView-2 imagery for mapping leaf area index of mangrove forest. *International Journal of Applied Earth Observation and Geoinformation*. <https://doi.org/10.1016/j.jag.2017.05.002>.
- Tilman, David, Christian Balzer, Jason Hill, and Belinda L. Befort. 2011. Global food demand and the sustainable intensification of agriculture. *Proceedings of the National Academy of Sciences of the United States of America*. <https://doi.org/10.1073/pnas.1116437108>.
- Tomczak, Maciej, Ewa Tomczak, Pawel Kleka, and Robert Lew. 2014. Using power analysis to estimate appropriate sample size. *Trends in Sport Sciences*.
- Torres-Sánchez, J., F. López-Granados, and J. M. Peña. 2015. An automatic object-based method for optimal thresholding in UAV images: Application for vegetation detection in herbaceous crops. *Computers and Electronics in Agriculture* 114: 43–52. <https://doi.org/10.1016/j.compag.2015.03.019>.
- Torresan, Chiara, Andrea Berton, Federico Carotenuto, Salvatore Filippo Di Gennaro, Beniamino Gioli, Alessandro Matese, Franco Miglietta, Carolina Vagnoli, Alessandro Zaldei, and Luke Wallace. 2017. Forestry applications of UAVs in Europe: a review. *International Journal of Remote Sensing*. <https://doi.org/10.1080/01431161.2016.1252477>.
- Tucker, Compton J. 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment* 8: 127–150. [https://doi.org/10.1016/0034-4257\(79\)90013-0](https://doi.org/10.1016/0034-4257(79)90013-0).
- Turner, Darren, Arko Lucieer, and Christopher Watson. 2012. An automated technique for generating georectified mosaics from ultra-high resolution Unmanned Aerial Vehicle (UAV) imagery, based on Structure from Motion (SfM) point clouds. *Remote Sensing*. <https://doi.org/10.3390/rs4051392>.
- Ubbens, Jordan, Mikolaj Cieslak, Przemyslaw Prusinkiewicz, and Ian Stavness. 2018. The use of plant models in deep learning: An application to leaf counting in rosette plants. *Plant Methods*. <https://doi.org/10.1186/s13007-018-0273-z>.
- United Nations. 2015. World Population Prospects: The 2015 Revision, Key Findings and Advance Tables. Working Paper No. ESA/P/WP.241. *United Nations, Department of Economic and Social Affairs, Population Division*.
- Urban, R., T. Reindl, and J. Brouček. 2020. Testing of drone DJI Phantom 4 RTK accuracy. In *Advances and Trends in Geodesy, Cartography and Geoinformatics II*. <https://doi.org/10.1201/9780429327025-16>.
- Valério, Igor Pirez, Fernando Irajá Félix De Carvalho, Antônio Costa De Oliveira, Giovanni Benin, Velci Queiroz De Souza, Amauri De Almeida Machado, Ivandro Bertan, Cyrano Cardoso Busato, Gustavo Da Silveira, and Daniel Andrei Rob Fonseca. 2009. Seeding density in wheat genotypes as a function of tillering potential. *Scientia Agricola*. <https://doi.org/10.1590/s0103-90162009000100004>.
- Vargas, Juan Quirós, Juliane Bendig, Alasdair Mac Arthur, Andreas Burkart, Tommaso Julitta, Kadmiel Maseyk, Rick Thomas, et al. 2020. Unmanned Aerial Systems (UAS)-Based Methods for Solar Induced Chlorophyll Fluorescence (SIF) Retrieval with Non-Imaging Spectrometers: State of the Art. *Remote Sensing*. <https://doi.org/10.3390/RS12101624>.
- Verrelst, Jochem, Gustau Camps-Valls, Jordi Muñoz-Marí, Juan Pablo Rivera, Frank Veroustraete, Jan G.P.W. Clevers, and José Moreno. 2015. Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties - A review. *ISPRS Journal of Photogrammetry and Remote Sensing*. <https://doi.org/10.1016/j.isprsjprs.2015.05.005>.
- Wahab, Ibrahim, Ola Hall, and Magnus Jirström. 2018. Remote sensing of yields: Application of UAV imagery-derived ndvi for estimating maize vigor and yields in

- complex farming systems in Sub-Saharan Africa. *Drones*.  
<https://doi.org/10.3390/drones2030028>.
- Walsh, Olga S., Sanaz Shafian, Juliet M. Marshall, Chad Jackson, Jordan R. McClintick-Chess, Steven M. Blanscet, Kristin Swoboda, Craig Thompson, Kelli M. Belmont, and Willow L. Walsh. 2018. Assessment of UAV Based Vegetation Indices for Nitrogen Concentration Estimation in Spring Wheat. *Advances in Remote Sensing*.  
<https://doi.org/10.4236/ars.2018.72006>.
- Ward, Sean, Jordon Hensler, Bilal Alsalam, and Luis Felipe Gonzalez. 2016. Autonomous UAVs wildlife detection using thermal imaging, predictive navigation and computer vision. In *IEEE Aerospace Conference Proceedings*.  
<https://doi.org/10.1109/AERO.2016.7500671>.
- Weibel, R O, and J W Pendleton. 1964. Effect of Artificial Lodging on Winter Wheat Grain Yield and Quality (AJ). *Agronomy Journal* 48: 452–455.  
<https://doi.org/10.2134/agronj1964.00021962005600050013x>.
- Weiss, M., F. Jacob, and G. Duveiller. 2020. Remote sensing for agricultural applications: A meta-review. *Remote Sensing of Environment*.  
<https://doi.org/10.1016/j.rse.2019.111402>.
- Westoby, M J, J Brasington, N F Glasser, M J Hambrey, and J M Reynolds. 2012. ‘Structure- from- Motion’ photogrammetry: A low- cost, effective tool for geoscience applications. *Geomorphology* 179: 300–314.  
<https://doi.org/10.1016/j.geomorph.2012.08.021>.
- Whaley, J. M., D. L. Sparkes, M. J. Foulkes, J. H. Spink, T. Semere, and R. K. Scott. 2000. The physiological response of winter wheat to reductions in plant density. *Annals of Applied Biology*. <https://doi.org/10.1111/j.1744-7348.2000.tb00048.x>.
- Willmott, Cort J., and Kenji Matsuura. 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*. <https://doi.org/10.3354/cr030079>.
- Winter, S. R., J. T. Musick, and K. B. Porter. 1988. Evaluation of Screening Techniques for Breeding Drought-Resistant Winter Wheat. *Crop Science*.  
<https://doi.org/10.2135/cropsci1988.0011183x002800030018x>.
- Xie, Chuanqi, and Ce Yang. 2020. A review on plant high-throughput phenotyping traits using UAV-based sensors. *Computers and Electronics in Agriculture*.  
<https://doi.org/10.1016/j.compag.2020.105731>.
- Yang, Guijun, Jiangang Liu, Chunjiang Zhao, Zhenhong Li, Yanbo Huang, Haiyang Yu, Bo Xu, et al. 2017a. Unmanned aerial vehicle remote sensing for field-based crop phenotyping: Current status and perspectives. *Frontiers in Plant Science*.  
<https://doi.org/10.3389/fpls.2017.01111>.
- Yang, Ming Der, Kai Siang Huang, Yi Hsuan Kuo, Hui Ping Tsai, and Liang Mao Lin. 2017b. Spatial and spectral hybrid image classification for rice lodging assessment through UAV imagery. *Remote Sensing* 9. <https://doi.org/10.3390/rs9060583>.
- Zhang, Chunhua, and John M. Kovacs. 2012. The application of small unmanned aerial systems for precision agriculture: A review. *Precision Agriculture* 13: 693–712.  
<https://doi.org/10.1007/s11119-012-9274-5>.
- Zhang, Heping, Neil C. Turner, Michael L. Poole, and Senthold Asseng. 2007. High ear number is key to achieving high wheat yields in the high-rainfall zone of south-western Australia. *Australian Journal of Agricultural Research*.  
<https://doi.org/10.1071/AR05170>.
- Zhang, Mingwei, Hui Wang, Yuan Yi, Jinfeng Ding, Min Zhu, Chunyan Li, Wenshan Guo, Chaonian Feng, and Xinkai Zhu. 2017. Effect of nitrogen levels and nitrogen ratios on lodging resistance and yield potential of winter wheat (*Triticum aestivum* L.). *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0187543>.

- Zhou, Chengquan, Dong Liang, Xiaodong Yang, Bo Xu, and Guijun Yang. 2018a. Recognition of wheat spike from field based phenotype platform using multi-sensor fusion and improved maximum entropy segmentation algorithms. *Remote Sensing*. <https://doi.org/10.3390/rs10020246>.
- Zhou, Chengquan, Dong Liang, Xiaodong Yang, Hao Yang, Jibo Yue, and Guijun Yang. 2018b. Wheat ears counting in field conditions based on multi-feature optimization and TWSVM. *Frontiers in Plant Science*. <https://doi.org/10.3389/fpls.2018.01024>.
- Zhu, Yanjun, Zhiguo Cao, Hao Lu, Yanan Li, and Yang Xiao. 2016. In-field automatic observation of wheat heading stage using computer vision. *Biosystems Engineering*. <https://doi.org/10.1016/j.biosystemseng.2015.12.015>.

## 10. Supplementary materials

**Table S1** Summary of the variance analysis of barley based on the MAE for different parameters (VI, BBCH, GSD). B: barley; VI: vegetation index; BBCH: Biologische Bundesanstalt, Bundessortenamt und Chemische Industrie; GSD: Ground sampling distance; \*\*\* significance level < 0.0001.

	Df	Sum Sq	Mean Sq	F Value	Pr(>F)
B\$VI	1	5051	5051	260.27	
B\$BBCH	3	15117	5039	259.64	< 2.20e-16 ***
B\$GSD	1	3034.4	3034	156.36	
B\$VI:B\$BBCH	3	5158.4	1720	88.60	
B\$VI:B\$GSD	1	394.9	395	20.35	7.49e-06 ***
B\$BBCH:B\$GSD	3	11416	3805	196.08	< 2.20e-16 ***
B\$VI:B\$BBCH:B\$GSD	3	9836.7	3279	1.69E+02	
Residuals	752	14594	19.4		

**Table S2** Results of the three-way ANOVA for barley based on the MAE. Multiple R-squared: 0.77; adjusted R-squared: 0.77; F-statistic: 171.8 on 15 and 752 DF; p-value: 2.20e-16.

	Estimate	Std. Error	t Value	Pr(> t )
(Intercept)	54.94	0.64	86.40	
B\$VI_NDVI	-26.38	0.90	-29.33	
B\$BBCH_13	-24.17	0.90	-26.88	
B\$BBCH_14	-21.02	0.90	-23.37	
B\$BBCH_15	-16.28	0.90	-18.10	
B\$GSD_High	-29.12	0.90	-32.39	
B\$VI_NDVI:B\$BBCH_13	22.18	1.27	17.44	
B\$VI_NDVI:B\$BBCH_14	30.61	1.27	24.07	
B\$VI_NDVI:B\$BBCH_15	26.47	1.27	20.81	< 2.00e-16 ***
B\$VI_NDVI:B\$GSD_High	24.62	1.27	19.36	
B\$BBCH_13:B\$GSD_High	23.87	1.27	18.77	
B\$BBCH_14:B\$GSD_High	39.53	1.27	31.08	
B\$BBCH_15:B\$GSD_High	31.46	1.27	24.74	
B\$VI_NDVI:B\$BBCH_13:B\$GSD_High	-18.98	1.80	-10.55	
B\$VI_NDVI:B\$BBCH_14:B\$GSD_High	-38.39	1.80	-21.34	
B\$VI_NDVI:B\$BBCH_15:B\$GSD_High	-29.64	1.80	-16.48	



**Table S3** Summary of the variance analysis of wheat based on MAE for different parameters (VI, BBCH, GSD). W: wheat.

	Df	Sum Sq	Mean Sq	F Value	Pr(>F)
W\$VI	1	25381	25381	860	
W\$BBCH	3	25269	8423	285	
W\$GSD	1	15276	15276	517	
W\$VI:W\$BBCH	3	5949	1983	67	< 2.2e-16 ***
W\$VI:W\$GSD	1	9349	9349	317	
W\$BBCH:W\$GSD	3	2874	958	32	
W\$VI:W\$BBCH:W\$GSD	3	9293	3098	105	
Residuals	752	22206	30		

**Table S4** Results of the three-way ANOVA for wheat based on MAE. Multiple R-squared: 0.81; adjusted R-squared: 0.81; F-statistic: 210.8 on 15 and 752 D; p-value: < 2.20e-16; \*\*\* significance level < 0.0001.

	Estimate	Std. Error	t Value	Pr(> t )
(Intercept)	67.47	0.78	86.02	< 2.00e-16 ***
W\$VI_NDVI	-25.79	1.11	-23.25	< 2.00e-16 ***
2	-5.09	1.11	-4.59	5.15e-06 ***
W\$BBCH_13	-32.76	1.11	-29.53	< 2.00e-16 ***
W\$BBCH_14	-21.01	1.11	-18.94	< 2.00e-16 ***
W\$GSD_High	-27.30	1.11	-24.62	< 2.00e-16 ***
W\$VI_NDVI:W\$BBCH_12	-7.76	1.57	-4.95	9.38e-07 ***
W\$VI_NDVI:W\$BBCH_13	23.63	1.57	15.06	< 2.00e-16 ***
W\$VI_NDVI:W\$BBCH_14	13.39	1.57	8.54	< 2.00e-16 ***
W\$VI_NDVI:W\$GSD_High	25.11	1.57	16.00	< 2.00e-16 ***
W\$BBCH_12:W\$GSD_High	2.57	1.57	1.64	0.101
W\$BBCH_13:W\$GSD_High	23.88	1.57	15.22	< 2.00e-16 ***
W\$BBCH_14:W\$GSD_High	19.17	1.57	12.22	< 2.00e-16 ***
W\$VI_NDVI:W\$BBCH_12:W\$GSD_High	5.17	2.22	2.33	0.02 *
W\$VI_NDVI:W\$BBCH_13:W\$GSD_High	-26.67	2.22	-12.02	< 2.00e-16 ***
W\$VI_NDVI:W\$BBCH_14:W\$GSD_High	-23.10	2.22	-10.41	< 2.00e-16 ***

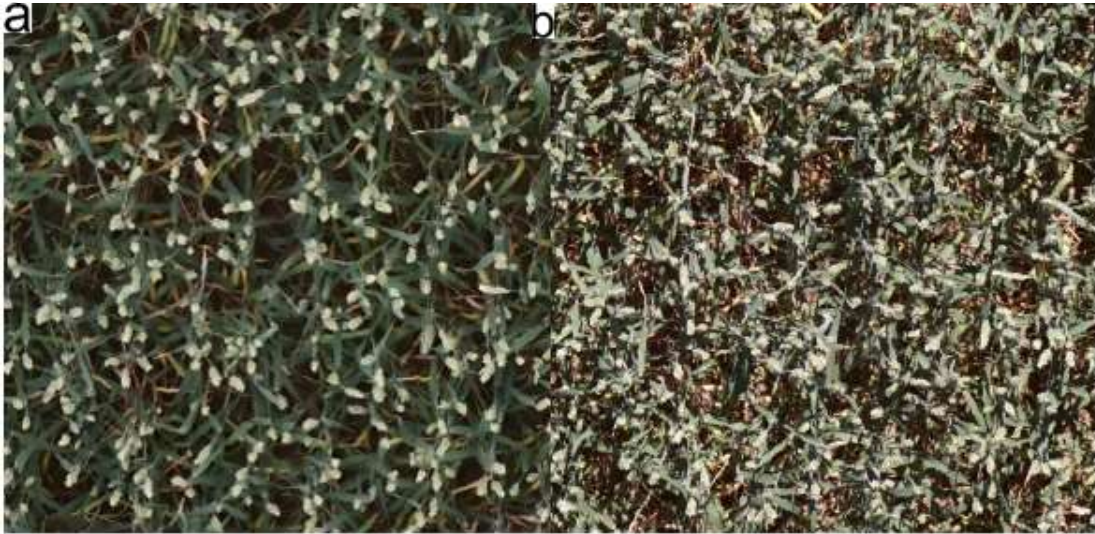
**Table S5** Lower and upper thresholds of extended CIELAB color space identified for each BBCH stage, genotype and GSD for spike segmentation. L: Lower threshold; U: Upper thresh

BBCH	Color dimension	Tybalt						Kadrilj						Quintus					
		1.4 mm GSD		2.0 mm GSD		2.8 mm GSD		1.4 mm GSD		2.0 mm GSD		2.8 mm GSD		1.4 mm GSD		2.0 mm GSD		2.8 mm GSD	
		L	U	L	U	L	U	L	U	L	U	L	U	L	U	L	U	L	U
61	L*	165	255	155	255	145	255	165	255	155	255	145	255	160	255	150	255	130	255
	a*	115	125	115	125	115	125	115	125	115	125	115	125	0	255	0	255	0	255
	b*	0	255	0	255	0	255	0	255	0	255	0	255	0	255	0	255	0	255
	C	0	178	0	178	0	178	0	178	0	178	0	178	176	255	176	255	176	255
	H	131	255	131	255	131	255	131	255	131	255	131	255	137	255	137	255	137	255
71	L*	160	255	165	255	170	255	160	255	177	255	170	255	140	255	145	255	140	255
	a*	0	134	0	134	0	134	0	132	0	131	0	131	0	130	0	126	0	127
	b*	0	255	0	255	0	255	0	255	0	255	0	255	0	255	0	255	0	255
	C	0	199	0	200	0	200	0	199	0	198	0	197	0	195	0	197	0	197
	H	133	255	131	255	130	255	132	255	130	255	131	255	134	255	138	255	138	255
77	L*	195	255	185	255	175	255	195	255	185	255	180	255	167	255	155	255	155	255
	a*	0	134	0	132	0	135	0	133	0	132	0	133	121	130	0	130	0	131
	b*	0	255	0	255	0	255	0	255	0	255	0	255	0	255	0	255	0	255
	C	0	200	0	201	0	202	0	199	0	201	0	202	0	199	0	200	0	199
	H	128	255	128	255	128	255	128	255	128	255	128	255	133	255	137	255	132	255
85	L*	180	255	178	255	190	255	180	255	178	255	190	255	180	255	180	255	185	255
	a*	0	141	0	141	0	139	0	141	0	141	0	139	0	141	0	141	0	135
	b*	0	255	0	255	0	255	0	255	0	255	0	255	0	255	0	255	0	255
	C	0	206	0	206	0	204	0	206	0	206	0	204	0	205	0	205	0	209
	H	120	255	120	255	120	255	120	255	120	255	120	255	120	255	120	255	120	255

**Table S6** Stepwise reduction of the calibration sample size from 21 to six reference measurements for different GSD and for best-performing models at BBCH stage 71 for Tybalt, BBCH Stage 77 for Kadrij and BBCH stage 85 for Quintus. Remaining samples were used to validate the performance of the calibrated models. Averaged statistical measures were computed across the empirical regression models randomly divided into calibration and validation data sets for each sample size. Different letters indicate significant differences between RMSE ( $p < 0.05$ ).

n (Calibration/ Validation)	GSD (cm)	Tybalt							Kad						
		Calibration			Validation				Calibration			Validation			
		S	I	R <sup>2</sup>	S	I	R <sup>2</sup>	RMSE	S	I	R <sup>2</sup>	S	I	R <sup>2</sup>	RMSE
21/11	1.4	0.021	-0.475	0.82	0.979	9.72	0.83	15.91a	0.026	-2.393	0.76	0.964	15.875	0.74	16.26a
16/16		0.022	-0.931	0.82	1.042	-16.206	0.83	16.5a	0.027	-3.001	0.77	0.884	43.539	0.71	17.04a
11/21		0.021	-0.801	0.82	0.994	9.118	0.83	16.73a	0.026	-2.554	0.76	0.849	74.129	0.73	17.6a
6/26		0.022	-1.126	0.81	1.253	-106.155	0.83	21.75b	0.027	-2.919	0.78	1.176	-76.073	0.74	20.56b
21/11	2	0.021	1.92	0.73	0.933	23.71	0.73	20.71b	0.026	-2.609	0.64	0.999	-4.849	0.63	22.45c
16/16		0.019	2.508	0.71	1.039	-10.881	0.73	21.91b	0.026	-2.618	0.65	1.048	-32.764	0.61	25.8cd
11/21		0.022	1.421	0.71	0.961	8.646	0.72	22.37b	0.027	-3.06	0.66	1.186	-87.276	0.61	27.94d
6/26		0.021	1.721	0.71	1.13	-44.832	0.71	27.19c	0.03	-4.233	0.65	1.381	-160.296	0.63	33.43e
21/11	2.8	0.016	1.941	0.65	0.874	41.436	0.63	26.74c	0.027	-2.881	0.53	0.859	50.734	0.54	25.38c
16/16		0.016	2.14	0.64	0.993	11.413	0.61	29.15cd	0.028	-3.524	0.56	0.94	35.291	0.5	28.58d
11/21		0.014	2.831	0.58	1.09	-24.16	0.61	32.03d	0.026	-2.557	0.54	1.203	-80.974	0.52	34.49e
6/26		0.015	2.267	0.61	1.088	-11.514	0.61	39.85e	0.023	-1.36	0.55	1.348	-138.132	0.53	39.79f

n (Calibration/ Validation)	GSD (cm)	Quintus						
		Calibration			Validation			
		S	I	R <sup>2</sup>	S	I	R <sup>2</sup>	RMSE
21/11	1.4	0.027	-3.390	0.81	0.988	4.527	0.81	16.71a
16/16		0.028	-3.575	0.81	1.055	-18.322	0.81	18.46a
11/21		0.029	-4.031	0.83	0.917	42.109	0.79	18.99a
6/26		0.028	-3.784	0.85	1.308	-119.014	0.80	25.16b
21/11	2	0.029	-3.942	0.44	1.166	-65.541	0.45	44.51c
16/16		0.030	-4.437	0.43	1.184	-66.679	0.48	45.96d
11/21		0.024	-1.660	0.38	1.370	-161.392	0.49	51.72e
6/26		0.033	-5.531	0.51	1.424	-184.752	0.48	55.6e
21/11	2.8	0.032	-5.526	0.36	1.120	-35.384	0.35	53.58e
16/16		0.029	-4.449	0.32	1.256	-118.956	0.41	56.49f
11/21		0.032	-5.732	0.38	1.269	-85.854	0.37	62.21f
6/26		0.029	-4.209	0.39	1.505	-195.597	0.35	71.59g



**Figure S1** Appearance of the same area in an orthomosaic, influenced by illumination conditions at BBCH stage 71 for Tybalt. **(a)** Data acquired under cloudy illumination conditions and **(b)** data acquired under constant solar illumination conditions, which do not allow for a good visual identification of the spikes, even with the naked eye.

## 11. Acknowledgements

Zunächst einmal möchte ich meinem Doktorvater, Uwe Rascher, meinen Dank aussprechen. Mit Deiner Persönlichkeit, Deinen Inspirationen und Deinem Enthusiasmus, den Du bereits im Studium versprüht hast, hast Du einen wesentlichen Anteil an meinem Werdegang genommen. Die Fähigkeit, Kritik so zu verpacken, dass man aus einer herausfordernden Situation neue Motivation schöpft, bewundere und schätze ich selbst nach den ganzen Jahren unserer Zusammenarbeit wie am ersten Tag.

Zusätzlich möchte ich ebenfalls Dir Bastian Siegmann danken, dass Du Dir über die Jahre hinweg stets versucht hast die Zeit für eine gute Betreuung zu nehmen. Ich weiß die Unterstützung während der Zeit sehr zu schätzen und dass du stets ein offenes Ohr für mich hattest. Die interessanten und lustigen Tagungen und Fortbildungen mit Dir werden in Erinnerung bleiben! Meinen zusätzlichen Dank möchte ich Johannes Postma widmen, der sich im Laufe der Promotion dazu bereit erklärt hat, dem Betreuerteam beizuwohnen. Neben der hohen Qualität des Feedbacks, war Deine spezielle Sichtweise und Dein Input ein Mehrwert, was mich inspiriert und motiviert hat.

Ein großer Dank gilt natürlich meiner Perla, Maren Mauel und meiner Familie insbesondere meiner Mutter Linda, meinem Vater Uwe, meiner aller liebsten Schwester Alica und meinem Lieblingsbruder Dustin. Ihr habt den Konjunkturzyklus, der zweifellos mit einer Promotion verbunden ist, mit mir durchlebt und es ist fast überflüssig zu erwähnen, dass ich ohne Eure Unterstützung diese Zeilen hier nicht schreiben könnte! Das Gleiche betrifft neben meinem Lieblingsnachbarn auch meine Weggefährten in Jülich, allen voran Felix Frimpong, Ines Munoz-Fernandez, Beat Keller, Philipp Norf, Vera Krieger, Erekle Chakhvashvili, Dzhaner Emin, Sascha Heinemann, Moritz Nabel und Ivan BARBOSA! Ich werde die tiefgründigen Gespräche mit Dir sehr vermissen.

Vergessen werden darf in diesem Abschnitt definitiv nicht Andreas Burkart! Danke, dass Du Dich bereits im Studium meiner angenommen und mich in die zur damaligen Zeit noch mysteriöse Welt der Drohnen eingeführt hast. Es erfüllt mich mit Stolz, dass Du mich mit dazu auserkoren hast, dein "Erbe" in Jülich anzutreten. Ich hoffe, die eine oder andere Zeile der Promotion stimmt dich freudig. Den Weg zur dunklen Seite der Macht bin ich jetzt ebenfalls mit dir gegangen. Abschließen möchte ich meine Promotion letztendlich mit:

*Wenn's einfach wäre, würd's jeder machen*

*If it were easy, everyone would do it*

Tschau, Tschüss und Byebye aus der Eifel!