LEONARDO VOLPATO

HIGH-THROUGHPUT PHENOTYPING FOR SOYBEAN PLANT MATURITY DATE AND WHEAT PLANT HEIGHT USING UNMANNED AERIAL SYSTEM

Thesis submitted to the Breeding and Genetics Graduate Program of the Universidade Federal de Viçosa, in partial fulfillment of the requirements to obtain the degree of *Doctor Scientiae*.

Adviser: Aluízio Borém

Co-advisers: Felipe Silva Lopes

Francelino Augusto Rodrigues Junior

VIÇOSA - MINAS GERAIS

2020

Ficha catalográfica elaborada pela Biblioteca Central da Universidade Federal de Viçosa - Campus Viçosa

T
V931h
2020

Volpato, Leonardo, 1992-
High-throughput phenotyping for soybean plant maturity date and wheat plant height using unmanned aerial system / Leonardo Volpato. – Viçosa, MG, 2020.
93 f. : il. (algumas color.) ; 29 cm.

Inclui apêndices.
Orientador: Aluízio Borém de Oliveira.
Tese (doutorado) - Universidade Federal de Viçosa.
Inclui bibliografia.

1. Melhoramento genético. 2. *Glycine max*. 3. *Triticum* sp. 4. Drone. 5. Imagens fotográficas. 6. Sensoriamento remoto. I. Universidade Federal de Viçosa. Departamento de Agronomia. Programa de Pós-Graduação em Genética e Melhoramento. II. Título.

CDD 22.ed. 631.52

LEONARDO VOLPATO

HIGH-THROUGHPUT PHENOTYPING FOR SOYBEAN PLANT MATURITY DATE AND WHEAT PLANT HEIGHT USING UNMANNED AERIAL SYSTEM

Thesis submitted to the Breeding and Genetics Graduate Program of the Universidade Federal de Viçosa, in partial fulfillment of the requirements to obtain the degree of *Doctor Scientiae*.

APPROVED: October 20, 2020.

Assent:

Leonardo Volpato
Author

Aluízio Borém de Oliveira
Adviser

To the divine Master, friends and professors,

and my family,

for their unconditional commitment and support

to make my dreams come true.

ACKNOWLEDGMENTS

To God, for his perfection and greatness, to spiritual friends for granting me serenity and courage in moments of insecurity and uncertainty.

To Federal University of Viçosa, through the Graduate Program in Genetics and Breeding and the Department of Agronomist for the opportunity to grant me with Agronomist, M.Sc., and Ph.D. degree.

To my parents, Cleide Aparecida Vança Volpato and Vanderlei Volpato, for their love, affection, support, and example of life.

To my sister, Miriane Aparecida Vança Volpato, for her love, affection, support, patience, and encouragement.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

To the Brazilian Federal Foundation for Support and Evaluation of Graduate Education - CAPES, for granting a scholarship abroad.

To my academic advisor, Prof. Aluízio Borém, for his guidance, learning, and friendship. I sincerely appreciate the extensive timely feedback and guidance on research and technical communication.

To Prof. Felipe Lopes da Silva, for giving me a chance to work with and learn from him in the Soybean Program.

To the scientists Dr. Francelino Rodrigues Jr. and Dr. Aaron Lorenz for their guidance and motivation during my exchange period of academic training at CIMMYT and UMN.

To the graduate students' degree of the Oilseeds Laboratory for their support and to the opportunity to socialize and interact with them.

To the Soybean Program students, João Marques Soares Ferreira, Júlia de Araújo Rodrigues Nascimento, and Lucas de Amaral Silvestre for their friendship and unconditional supporting to trials conducting.

To the employees from the Department of Agronomist for support in the execution of field works.

Finally, my acknowledgment of everyone who contributed directly or indirectly to performing this work.

THANK YOU SO MUCH!!!

BIOGRAPHY

Leonardo Volpato, son of Vanderlei Volpato and Cleide Aparecida Vança Volpato, was born on January 4, 1992 at Araguari city - MG, located in the Triângulo Mineiro.

In 2010, he completed high school along with a technical degree in agriculture and livestock production at the Federal Institute of Education and Technology of the Triangulo Mineiro - Campus Uberlândia.

In 2015, he received a bachelor's degree in 2015 in Agronomy at the Federal University of Viçosa (UFV), Viçosa-MG.

In March 2015, he joined the Master's degree in Genetics and Breeding at the Federal University of Viçosa (UFV), Viçosa-MG, concluding the title in 2016.

In August 2016, he started a Ph.D. degree in Genetics and Breeding at the Federal University of Viçosa (UFV), Viçosa-MG, concluding the title in 2020.

From September 2018 to July 2019, Volpato made a training as Visiting Scholar at International Maize and Wheat Improvement Center - CIMMYT, MEX.

From August 2019 to March 2020, Volpato participated as Research Scholar at the University of Minnesota - UMN, US.

ABSTRACT

VOLPATO, Leonardo, D.Sc., Universidade Federal de Viçosa, October, 2020. **High-throughput phenotyping for soybean plant maturity date and wheat plant height using unmanned aerial system.** Adviser: Aluízio Borém. Co-advisers: Felipe Lopes da Silva and Francelino Rodrigues Junior.

Improvements in agronomical practices and plant breeding are paramount responses to the present and future challenges imposed by biotic stress and abiotic food production factors. On what concerns breeding currently, constraints in-field phenotyping capability limit our ability to dissect quantitative traits' genetics, especially those related to yield and environmental stress tolerance. Advances in phenotyping technology are critical to ensure that crops' genetic improvement meets future global demands for food and fuel. Progress in sensors, aeronautics, and high-performance computing combine with robust statistical analysis is paving this way. Throughout history, improving traits of interest depends on the ability to quantify phenotypes across genotypes replicated over multiple environments. Therefore, potentially valuable traits may have been neglected due to costly phenotyping and technological limitations. Field-based high throughput phenotyping (HTP) platforms will combine non-invasive remote sensing methods and automated environmental data collection. The current work aims to handle phenotypic data establishing UAS-based HTP platforms to improve data quality and provide a feasibility pipeline implementation applied to plant breeding programs. More specifically, chapter one deals with spring wheat (*Triticum* sp.), while chapter two, with the methodologies used in soybean breeding programs [*Glycine max* (L.) Merrill]. Plant height from two spring wheat breeding cycles was assessed using plot-level information from aerial imaging-derived 3D crop surface models at different phenological stages. From five contrasting environments, soybean maturity date with 53 experimental trials was collected by time-course images using RGB cameras equipped with low-cost drones. These studies showed that the phenotypic data obtained from a time-series imagery estimation via the UAS-based HTP platform could improve the collected data field's efficiency and quality and scalable to tens-of-thousands of plots into a modern plant breeding program.

Keywords: Drone imagery. RGB camera. Data collect. Crop surface model. Plant maturity. *Glycine max*. *Triticum* sp.

RESUMO

VOLPATO, Leonardo, D.Sc., Universidade Federal de Viçosa, outubro de 2020. **Fenotipagem de alto rendimento para data de maturidade em soja e altura planta em trigo via sistema aéreo não tripulado.** Orientador: Aluízio Borém. Coorientadores: Felipe Lopes da Silva e Francelino Rodrigues Junior.

Avanços na tecnologia de fenotipagem são fundamentais para garantir que o melhoramento genético das culturas, atenda às futuras demandas globais de alimentos e combustíveis. No que diz respeito ao melhoramento, atualmente, restrições na capacidade de fenotipagem de campo limitam a capacidade de dissecar características genéticas, principalmente, de natureza quantitativa, especialmente aquelas relacionadas a produção e à tolerância ao estresse ambiental. As plataformas de fenotipagem de alto rendimento (HTP) a nível de campo tende a combinar métodos de sensoriamento remoto não invasivos, juntamente com a coleta automatizada de dados ambientais. O objetivo do presente trabalho é apresentar alternativas para lidar com os dados fenotípicos de campo estabelecendo plataformas para HTP via drones (sistemas aéreos não tripuláveis - UAS) para coleta e precisão na qualidade dos dados de campo, assim como, implementação de um esquema de trabalho a ser viabilizado ao programa de melhoramento de plantas. Mais especificamente, o capítulo um, trata a respeito da cultura do trigo (*Triticum* sp.), enquanto o capítulo dois, as metodologias utilizadas foram destinadas aos programas de melhoramento genético de soja [*Glycine max* (L.) Merrill]. A altura de plantas de dois ciclos de melhoramento de trigo de primavera foi avaliada utilizando informações a nível de parcela oriundos de imagens areias derivadas do modelo 3D da superfície da cultura, em diferentes estádios fenológicos. A data de maturidade das parcelas de soja de cinco ambientes contrastantes com 53 ensaios experimentais ao total foi coletada por imagens ao longo do tempo via drones de baixo custo equipados com câmeras RGB. Os resultados sugeriram que os dados fenotípicos obtidos a partir de imagens aéreas oriundas de séries temporais via plataformas de HTP, podem melhorar a eficiência e a qualidade na coleta de dados a nível de campo, com potencial de ser escalável para dezenas de milhares de parcelas dentro de um programa moderno de melhoramento de plantas.

Palavras-chave: Imagens de drones. Câmera RGB. Coleta de dados. Modelo de superfície de cultura. Maturidade de plantas. *Glycine max* (L.) Merrill. *Triticum* sp.

SUMMARY

GENERAL INTRODUCTION

One of the most considerable challenges of the 21st century will be to increase crop yield to meet growing demands for food, clothing, and fuel due to both the increasing human population and its affluence. The most sustainable way to meet these demands is to develop and supply highly productive cultivars to farmers (Hickey et al., 2019). The recent biotech revolution is impacting the power and efficiency of the genetic improvement process of crops by increasing the capabilities of scientists to analyze large populations of plants with abundant genetic markers quickly, also aiming to fully sequence the genome of a large number of crops in breeding programs (Thomson et al., 2014; Mir et al., 2019). However, genetic analysis has its highest value when it can be associated with measurements of phenotyping traits arising from the object of study, that means high-throughput accurate acquisition and analysis of multidimensional phenotypes on an organism-wide scale through crop development (Yang et al., 2020). Although our ability to analyze DNA has been increasing exponentially, phenotyping studies have not equivalent improvement at the same level. Field phenotypic data acquisition is still a bottleneck restricting crop breeding and functional genomics studies (Furbank and Tester, 2011; Hu et al., 2020), as well as to validate new approaches to high-throughput phenotyping (HTP) (Reynolds et al., 2020; Yang et al., 2020). Traditional crop phenotyping methods are labor-intensive, time-consuming, subjective, and frequently plant destructive (Chen et al., 2014; Langre et al., 2019). However, in recent years, advanced sensor, machine vision, telemetry, robotics, and automation technology have been widely adopted across the agri-food industry and digital agriculture to augment efficiency, with applications ranging from phenotypic traits identification and develop bioinformatics technologies, as well as product quality assessment (Zhao et al., 2019).

Most modern breeders still obtain visual ratings and assessments of either individual plots or plants to estimate critical agronomic traits, just as they did decades ago. Actual data notes and subjective criterium by human ratings to measurements breeding fields might generate inaccuracies on the phenotyping collection data, limiting the visual classifications performed by scientists, which are carried out at the ground level by the breeder expertise. Besides, it is challenging to have visual ratings of tens-of-thousands of plots evaluated in modern breeding programs using the current techniques within a short period consuming lower cost and less labor-intensive (Haghighattalab et al., 2016; Zhao et al., 2019).

The introduction of HTP approaches can significantly improve the phenotyping standards for agronomic traits, contributing to understanding better their genetic basis and diversity, and the environmental influences throughout the crop's development cycle (Reynolds et al., 2020). Thus, the genotypes can be characterized using high dimensional marker information and high dimensional phenotypic information (Jarquin et al., 2018). The fast development of sensors and less expensive technologies has enabled unmanned aerial systems (UASs) use, such as drones, in the recent years. At the same pace, the development of image and data analyses algorithms along with best performance computers have enabled a broad range of possibilities for aerial HTP purposes of plant traits such as plant height, plant maturity, stand count, vegetation indices, among other target traits (Torres-Sánchez et al., 2013; Peña et al., 2015; Moreira et al., 2019; Matias et al., 2020; Morales et al., 2020). Recent advances in remote sensing using drones with sensors working in the visible (RGB - red, green, and blue) wavelength made possible to create high-throughput, cost-effective, and accurate quantitative phenotyping datasets in crop breeding programs (Yang et al., 2017).

Therefore, the drone-based HTP platform could be customized as a quick and at low-cost approach for agronomic trait evaluation to improve crop breeding programs' efficiency. However, field phenotyping must go hand-in-hand with methods to characterize and control field site variations for improving repeatability. Thus, a breeding workflow has to adopt appropriate experimental designs, select the right traits, and finally, proper integration of heterogeneous data sets, analysis, and applications, including prediction models (Araus and Cairns, 2014). Shortly, what will pave the way for the adoption of field HTP is to efficiently integrate all system components, including more user-friendly data management combined with data gathering and processing and suitable statistical methods.

Thus, the present work was carried out to study modern tools to collect and analyze phenotypic image data at the field level applied to plant breeding programs. In this scenario, the present study proposes to develop and validate a practical HTP pipeline applied to wheat and soybean breeding programs to estimate plant height and plant maturity, respectively, using RGB cameras mounted in a low-cost drone-based, as well as investigate the practical environmental influence under the remote sensing imagery approaches.

REFERENCES

Araus, J., and Cairns, J. (2014). Field high-throughput phenotyping: the new crop breeding frontier. *Trends Plant Sci.* 19, 52–61.

Chen, D., Neumann, K., Friedel, S., Kilian, B., Chen, M., Altmann, T., et al. (2014). Dissecting the phenotypic components of crop plant growthand drought responses based on high-throughput image analysis w open. *Plant Cell* 26, 4636–4655. doi:10.1105/tpc.114.129601.

Langre, E., Penalver, O., Hémon, P., Frachisse, J.-M., Bogeat-Triboulot, M.-B., Niez, B., et al. (2019). Nondestructive and Fast Vibration Phenotyping of Plants. *Plant Phenomics* 2019, 1–10. doi:10.34133/2019/6379693.

Furbank, R. T., and Tester, M. (2011). Phenomics - technologies to relieve the phenotyping bottleneck. *Trends Plant Sci.* 16, 635–644. doi:10.1016/j.tplants.2011.09.005.

Haghighattalab, A., González Pérez, L., Mondal, S., Singh, D., Schinstock, D., Rutkoski, J., et al. (2016). Application of unmanned aerial systems for high throughput phenotyping of large wheat breeding nurseries. *Plant Methods* 12. doi:10.1186/s13007-016-0134-6.

Hickey, L. T., N. Hafeez, A., Robinson, H., Jackson, S. A., Leal-Bertioli, S. C. M., Tester, M., et al. (2019). Breeding crops to feed 10 billion. *Nat. Biotechnol.* doi:10.1038/s41587-019-0152-9.

Hu, Y., Knapp, S., Schmidhalter, U., Mishra, P., Asaari, M. S. M., Herrero-Langreo, A., et al. (2020). Yield Prediction By Machine Learning From Uas-Based Multi-Sensor Data Fusion In Soybean. *Adv. Biochem. Eng. Biotechnol.* 16:78, 1–20. doi:10.1016/j.tplants.2018.11.007.

Jarquin, D., Howard, R., Xavier, A., and Choudhury, S. Das (2018). Increasing predictive ability by modeling interactions between environments, genotype and canopy coverage image data for soybeans. *Agronomy* 8. doi:10.3390/agronomy8040051.

Maes, W.H., and K. Steppe. 2019. Perspectives for Remote Sensing with Unmanned Aerial Vehicles in Precision Agriculture. Trends Plant Sci. 24(2): 152–164. doi: 10.1016/j.tplants.2018.11.007.

Matias, F. I., Caraza-harter, M. V, and Endelman, J. B. (2020). FIELDimageR : An R package to analyze orthomosaic images from agricultural field trials. 1–6. doi:10.1002/ppj2.20005.

Mir, R. R., Reynolds, M., Pinto, F., Khan, M. A., and Bhat, M. A. (2019). High-throughput phenotyping for crop improvement in the genomics era. *Plant Sci.* 282, 60–72. doi:10.1016/j.plantsci.2019.01.007.

Morales, N., Mueller, L. A., Kaczmar, N. S., Robbins, K. R., Santantonio, N., and Gore, M. A. (2020). ImageBreed : Open-access plant breeding web – database for image-based phenotyping Field experiments. 1–7. doi:10.1002/ppj2.20004.

Moreira, F. F., Hearst, A. A., Cherkauer, K. A., and Rainey, K. M. (2019). Improving the efficiency of soybean breeding with high-throughput canopy phenotyping. *Plant Methods* 15, 1–9. doi:10.1186/s13007-019-0519-4.

Peña, J. M., Torres-Sánchez, J., Serrano-Pérez, A., de Castro, A. I., and López-Granados, F. (2015). Quantifying efficacy and limits of unmanned aerial vehicle (UAV) technology for weed seedling detection as affected by sensor resolution. *Sensors (Switzerland)* 15, 5609–

5626. doi:10.3390/s150305609.

Reynolds, M., Chapman, S., Crespo-Herrera, L., Molero, G., Mondal, S., Pequeno, D. N. L., et al. (2020). Breeder Friendly Phenotyping. *Plant Sci.*, 110396. doi:10.1016/j.plantsci.2019.110396.

Thomson, M. J., Lawrence, T. R. N. and J. R., and Abstract (2014). High-Throughput SNP Genotyping to Accelerate Crop Improvement. *Adv. Biochem. Eng. Biotechnol.* 2, 195–212. doi:10.9787/pbb.2014.2.3.195.

Torres-Sánchez, J., López-Granados, F., De Castro, A. I., and Peña-Barragán, J. M. (2013). Configuration and Specifications of an Unmanned Aerial Vehicle (UAV) for Early Site Specific Weed Management. *PLoS One* 8. doi:10.1371/journal.pone.0058210.

Yang, G., Liu, J., Zhao, C., Li, Z., Huang, Y., Yu, H., et al. (2017). Unmanned Aerial Vehicle Remote Sensing for Field-Based Crop Phenotyping: Current Status and Perspectives. *Front. Plant Sci.* 8. doi:10.3389/fpls.2017.01111.

Yang, W., Feng, H., Zhang, X., Zhang, J., Doonan, J. H., Batchelor, W. D., et al. (2020). Crop Phenomics and High-Throughput Phenotyping: Past Decades, Current Challenges, and Future Perspectives. *Mol. Plant* 13, 187–214. doi:10.1016/j.molp.2020.01.008.

Yeyin, S., L. Neely, J. Thomasson, S. Murray, N. Pugh, et al. 2016. Unmanned Aerial Vehicles for High- Throughput Phenotyping and Agronomic. PLoS One c: 1–26. doi: 10.5061/dryad.65m87.

Zhao, C., Zhang, Y., Du, J., Guo, X., Wen, W., Gu, S., et al. (2019). Crop phenomics: Current status and perspectives. *Front. Plant Sci.* 10. doi:10.3389/fpls.2019.00714.

CHAPTER 1

HIGH THROUGHPUT FIELD PHENOTYPING FOR PLANT HEIGHT USING UAS-BASED RGB IMAGERY IN WHEAT BREEDING LINES: FEASIBILITY AND VALIDATION

VIÇOSA - MINAS GERAIS

2020

ABSTRACT

Plant height (PH) is an essential trait in the screening of most crops. While in crops such as wheat, medium stature helps reduce lodging, tall plants are preferred to increase total above-ground biomass. PH is an easy trait to measure manually, although it can be labor-intense depending on the number of plots. There is an increasing demand for alternative approaches to estimate PH in a higher throughput mode. Crop surface models (CSMs) derived from dense point clouds generated via aerial imagery could be used to estimate PH. This study evaluates PH estimation's feasibility at different phenological stages using plot-level information from aerial imaging-derived 3D CSM in wheat inbred lines during two consecutive years. Multi-temporal and high spatial resolution images were collected by a fixed-wing ($Plat_{FW}$) and a multi-rotor ($Plat_{MR}$) unmanned aerial system (UAS) platforms over two wheat populations (50 lines and 150 lines). The PHs were measured and compared at four growth stages (GS) using ground-truth measurements (PHground) and UAS-based estimates (PHaerial). The CSMs generated from the aerial imagery was validated using ground control points (GCPs) as fixed reference targets at different heights. The results show that PHs estimations using $Plat_{FW}$ were consistent with those obtained from $Plat_{MR}$, showing some slight differences due to image processing settings. The GCPs heights derived from CSM showed a high correlation and low error compared to their actual heights ($R^2 \geq 0.90$, RMSE $\leq$ 4 cm). The coefficient of determination (R^2) between PHground and PHaerial at different GS ranged from 0.35 to 0.88, and the RMSE from 0.39 to 4.02 cm for both platforms. In general, similar and higher heritability was obtained using PHaerial across different GS and years, and ranged according to the variability, and environmental error of the PHground observed (0.06 – 0.97). Finally, high Spearman rank correlations (0.47 – 0.91) and R^2 (0.63 – 0.95) of PHaerial adjusted and predicted values against PHground were also obtained. This study provides an example of UAS-based high-resolution RGB imagery use for obtaining time-series estimates of PHs of the crop surface, scalable to tens-of-thousands of plots suitable to be applied in plant wheat breeding trials.

Keywords: multi-temporal crop surface model; structure from motion; RGB camera; dense point cloud; drones; wheat breeding; adjusted and predicted genotypic values.

RESUMO

Altura de plantas (APs) é uma característica essencial utilizado para o melhoramento genético de várias culturas. Em culturas como o trigo, os genótipos anões são selecionados para reduzir o acamamento, enquanto em outras culturas agrícolas, plantas altas são visadas para a produção de bioenergia (sorgo, cana, etc.,). Modelos digitais de superfície vegetais de campo derivadas de nuvens de pontos densos, podem serem gerados a partir de imagens aéreas para estimar APs. O presente estudo objetivou-se mensurar APs por meio de informações em nível de parcela do modelo 3D de superfície da cultura (MSC). As imagens coletadas ao longo do tempo com resoluções e qualidade distintas, foram obtidas via plataformas de sistemas aéreos não tripuláveis (UAS) de asa fixa ($Plat_{FW}$) e multi-rotores ($Plat_{MR}$) de duas populações (50 e 150 linhas) de linhagens de trigo de primavera. As APs foram medidas e comparadas em quatro estágios distintos de crescimento (EC) durante dois ciclos de colheita, via medições coletadas diretamente no campo e estimadas por meio de UAS. Os MSCs gerados a partir das imagens aéreas foram validados utilizando pontos de controle no solo (PCSs) em diferentes alturas como alvos de referência fixos a longo dos ensaios. Os resultados mostram que as estimativas de APs usando ($Plat_{FW}$) foram consistentes com as obtidas via ($Plat_{MR}$), apresentando algumas pequenas diferenças devido às configurações de processamento e qualidade da imagem. As alturas dos PCSs derivados das plataformas UAS apresentaram alta correlação e baixo erro quando comparadas às suas alturas reais ($R^2 \geq 0{,}90$, RMSE $\leq$ 4 cm). O coeficiente de determinação das APs (R^2) entre as estimativas aéreas e as notas tomadas a nível de campo em diferentes EC variou entre 0,35 a 0,88, e o RMSE de 0,39 a 4,02 cm para ambas as plataformas. Alto R^2 (0,63 - 0,95) dos valores preditos também foram obtidos sob as interações GxA. Além disso, a herdabilidade genética para as APs estimadas via plataformas aéreas, variou de acordo com a variabilidade das medidas avaliadas a nível de campo e do erro ambiental (0,06 - 0,65). Finalmente, a seleção utilizando o valor de predição para APs é discutida. Este estudo fornece viável aplicabilidade para a estimação de APs baseadas em imagens de UAS coletadas ao longo do tempo, aplicável para dezenas de milhares de parcelas nas durante as avaliações elaboradas pelos programas de melhoramento genético.

Palavras-chave: modelo de superfície de cultura, estrutura de movimento, câmera RGB, nuvem de pontos densos, drones, imagens temporais, GCP e PPK para correções de geolocalização, valores genéticos ajustados e preditos.

1. Introduction

Wheat (*Triticum* sp.) is among the leading food crops, and it is grown in a range of environments and geographical areas. It is highly relevant to the human diet, given its protein quantity, quality, and variety of derived products (Shewry et al., 2016). Lately, wheat has become the most important source of dietary protein and the second most important source of calories (carbohydrates) for humans (Awika, 2011; Shewry et al., 2016). During the 2017/2018 season, wheat was sown in 11.7% of the world's arable land, corresponding to around 30% of world grain production (USDA, 2018). Wheat crops are vital for food security, supplying an affordable source of nutrition to a large portion of the global population, particularly millions of people with low-middle incomes (Peña-Bautista et al., 2017; Hickey et al., 2019), besides of being an essential crop for the composition of sustainable agricultural production systems (Moeller et al., 2014; Walters et al., 2016).

Led by Norman Borlaug, the wheat breeding contributed to the so-called "Green Revolution." Identification and introduction of major dwarfing or semi-dwarfing genes were significant advancements that enabled an actual grain yield increase in most environments by reducing plant height (PH) (Hedden, 2003; Reynolds and Borlaug, 2006; Würschum et al., 2015). Up to date, PH is one of the most critical and heritable traits in wheat breeding (Würschum et al., 2015). Breeders often phenotype for PH to select semi-dwarf cultivars in order to reduce lodging and improve grain yield and quality (Torres and Pietragalla, 2012; Chapman et al., 2014), as well as select suitable parental lines for hybrid breeding (Barmeier et al., 2016). Besides, PH contributes to biomass production probably associated with their ability to increase photosynthesis due to a better light interception and distribution through the canopy in taller plants (Song et al., 2013).

PH is commonly calculated by measuring the distance between the upper boundary of the main photosynthetic tissues (ignoring the awns) and the ground level using a simple metric ruler or a graduated stick (Torres and Pietragalla, 2012). Although simple, such assessment is costly, laborious, and prone to subjectivity, especially in extensive field trials of plant breeding (Holman et al., 2016; Hu et al., 2018). Additionally, manual ground measurements in the field are only feasible on a few plants per plot (Wilke et al., 2019), and therefore it can be biased by the lack of standardized criteria. Consequently, there is a need for faster, more accurate and continuous measurements of PH (Holman et al., 2016; Rebetzke et al., 2019), mainly because phenotyping under environmental field conditions is often considered as a bottleneck in plant breeding programs (Cobb et al., 2013; Yang et al., 2013). The temporal characterization of PH

could provide a better understanding of the mechanism of plant growth and underlying genetics effects by providing insight into environmental variables of the traits (Torres-Sánchez et al., 2013; Hassan et al., 2019a).

The introduction of high-throughput phenotyping (HTP) approaches into breeding schemes can significantly improve the phenotyping standards for agronomic traits, contributing to understanding better their genetic basis and diversity, and the environmental influences throughout the crop's development cycle (Reynolds et al., 2020). Non-destructive data collected via ground-based and aerial HTP techniques are highly desirable for applications in plant breeding since they can be used to assess target traits in large-scale field trials belonging to various plots dimensions (Mohan et al., 2017; Castro et al., 2018; Rodrigues et al., 2018; Zhang et al., 2018; Loladze et al., 2019). In particular, aerial HTP platforms have become favored over ground platforms as they are suitable for use in large breeding trials showing a good tradeoff between time, data accuracy, and resolution (Tattaris et al., 2016; Makanza et al., 2018; Gracia-Romero et al., 2019; Hassan et al., 2019a). The fast development of sensors and unmanned aerial system (UASs) observed in the recent years, together with the development of image and data analyses algorithms and improved computer capacities, has enabled a broad range of possibilities for aerial HTP purposes of plant traits such as PH, stand count, vegetation indices, among other agronomic traits (Torres-Sánchez et al., 2013; Peña et al., 2015; Matias et al., 2020; Morales et al., 2020; Moreira et al., 2020). These aerial imageries are used as a proxy for the characterization of quantitative plant traits. Recent advances in remote sensing using UAS with sensors working in the visible (RGB - red, green and blue bands) and/or near-infrared (NIR) have made it possible to create high-throughput, cost-effective and accurate quantitative phenotyping datasets in wheat breeding programs (Gracia-Romero et al., 2019; Singh et al., 2019; Reynolds et al., 2020) and for various other crops (Moreira et al., 2019; Tirado et al., 2019; Zhao et al., 2019; Radoglou-Grammatikis et al., 2020).

Digital color images (i.e., RGB) collected from UASs have been used for estimating PH in wheat through different settings and capabilities (Table 1). Feature matching and structure from motion (SfM) techniques applied to such imagery enable the generation of three-dimensional (3D) point clouds that can be used to reconstruct multi-temporal crop surface models (CSMs) from which PH can be estimated plot-wise (Westoby et al., 2012; Singh et al., 2016; Hassan et al., 2019b). SfM photogrammetry is a method that uses a set of overlapped images to generate high-resolution topographic 3D-reconstructions. Through automatic extraction of corresponding feature points, this method optimizes the 3D location based on images taken

from multiple perspectives, enabling a simple workflow (Westoby et al., 2012; James and Robson, 2014; Nex and Remondino, 2014).

Despite the advances in the use of SfM photogrammetry from UASs imagery for estimation of PH (PHaerial), there are several factors, such as image ground sampling distance (GSD) or weather conditions, that can potentially affect its performance and accuracy (Walter et al., 2015; Han et al., 2018; Lu et al., 2019). The implementation of an effective and low-cost workflow for PHaerial should consider these factors. In this view, an RGB camera can be deployed to PH estimation combine with adequate GSD and statistical aerial analysis under environmental conditions.

To the best of our knowledge, the image and data quality of ground-truth measurements (PHground) have not been adequately evaluated to assess their impact in PHaerial at individual plot growth stages (GS) in wheat breeding programs. Therefore, this study aims to validate PH derived from RGB imagery data and to understand the effect of data quality from different UAS platforms and PHground. The specific objectives are: (i) development of a semi-automated workflow for extraction, analysis, and evaluation of PHaerial at multiple GS; (ii) comparison of different UAS platforms used for PH estimations, and; (iii) assessment of the potential environmental issues associated between PHground and PHaerial. Finally, the genotypic values were obtained to investigate the PHaerial predictions.

Table 1. Summary of published studies on the estimation of plant height (PH) for wheat from RGB imagery acquired using unmanned aerial system (UASs)

Reference	GS	Platform - UAS	Camera / Sensor	GSD cm/p	Total number of plots	Plot size (m)	Pixel Extraction method	R^2 for PHground vs. Phaerial	RMSE for PHground vs. Phaerial (cm)
Holman et al. (2016)	GS30 to GS61	Cinestar Octocopter	Sony NEX 7 24.4 mgpx	1	300	9x3	99th percentile	0.52 - 0.99	1.5 - 9.9
Madec et al. (2017)	GS20 to GS69	Hexacopter	Sony ILCE-6000 digital$^{\$}$	1	1173	10x1.9	99.5th percentile	0.95 - 0.99	2.9 - 9.8
Hassan et al. (2019a)	GS41 and GS85	DJI Inspires 1 model T600	Sequoia 4.0 16 mgpx	2.5	600	1.3x1.3	90th and 99th percentile	0.8 - 0.96	5.75
Li et al. (2019)	GS21 to GS87	DJI Matrice 600 Pro	Zenmuse X5R RGB camera	0.5	170	3x0.23	Mean, median, 95th percentile, and standard deviation	-	-
Lu et al. (2019)	GS30 to GS69	DJI Phantom series	High-resolution digital RGB camera	1.66	36	6x5	Mean, median, standard deviation, coefficient of variation and 25th, 50th, 75th, and 95th percentiles	0.89	6

Cont. Table 1

Reference	GS	Platform - UAS	Camera / Sensor	GSD cm/p	Total number of plots	Plot size (m)	Pixel Extraction method	R^2 for PHground vs. Phaerial‡	RMSE for PHground vs. Phaerial (cm) ¶
Schirrmann et al. (2016)	GS41 to GS83	P-Y6, Hexapilots, Dresden (hexacopter)	Sony NEX 7 24 mgpx	1.2	20	1x1	90th percentile	0.76 - 0.92	6.0 - 15.0
Song and Wang, (2019)	GS31, GS65 and GS83	DJI Phantom 3	High-resolution digital RGB camera	1.5	15	2x2	Cuboid filter 3D classification	-	4.5 - 7.7
Yuan et al. (2018)	GS30 to GS87	Matrice 600 Pro (M600) and DJI Phantom 3 Pro	High-resolution digital RGB camera	0.47 – 0.67	100	1.5x1.524	89th and 100th percentiles	0.91	9
Yue et al. (2017)	GS30 and GS65	DJI S1000	UHD 185 Firefly - hyperspectral sensor#	1	48	6x8	Average of pixel values	0.69	19

GS, Growth stage; UAS, unmanned aerial system; GSD, ground sampling distances (cm/pixel). #Adapted sensor capturing 450–950-nm wavelengths. $30 mm and 19 mm focal length was used to equip the camera. ‡ Coefficient of determination (R^2) and ¶ Root mean square error (RMSE) between PHground and PHaerial, where the single numbers are the joining values obtained across GS by the respective authors.

2. Materials and methods

2.1. Plant material, site description, and data collection

The experiments were conducted over two spring wheat (*Triticum aestivum* L.) growing breeding cycles: 2016-2017 and 2017-2018 at the CIMMYT experimental station - Campo Experimental Norman E. Borlaug in Ciudad Obregon, northwestern Mexico (27◦20′N; 109◦54′ W; and 38 masl). Environmental and management details of this site are given in Sayre et al. (1997). Two spring wheat panels were studied under potential yield conditions: the high biomass association panel (HiBAP)-I and the HiBAP-II. Fifty inbreed lines were used for the validation in HiBAP-I during the 2016-2017 (Y17) and 2017-2018 (Y18) crop cycle, while the whole population of 150 lines was measured in HiBAP-II during the 2017-2018 (Y18). Both panels include representative lines derived from breeding and pre-breeding programs with a restricted range of maturity and height (Molero et al., 2019). The experimental design in both HiBAP panels consisted of an alpha-lattice design with two replicates and 30 incomplete blocks per replicate. The plots consisted of two beds in HiBAP-I Y17 and one bed in HiBAP I Y18 with two plant rows on the top of the beds for both trials. In HiBAP-II Y18, three replicates were evaluated in two beds plots. The beds in all three trials were 0.8 m wide, while the inter-row spacing within the bed and the space between beds were 0.24 m and 0.36 m, respectively. Plot length was 4 m for HiBAP I Y17 and HiBAP II Y18 and 2m for HiBAP I Y18.

Aerial (PHaerial - i.e., using UAS platforms) and ground-truth (PHground) plant height (PH) phenotyping were performed in the experiments during the following GSs: at 40 days after emergence (E+40), at booting (B), seven days after anthesis (A+7) and at physiological maturity (M). PHground was measured using a ruler when 50% of the plot reached the relevant GS as described by Torres and Pietragalla (2012). The flights between UASs were conducted on the same day or in one day of interval since the first plot reached a particular GS aiming optimal weather conditions (i.e., clear sky, around noon, and low wind speed). A summary of solar radiation and wind speed conditions during the entire flight campaigns for each platform used is given in Supplementary 1 Table. The average height was obtained from four random individual culms inside each plot (two in each bed), measuring the distance from the soil surface to the tip of the spike, excluding the awns, and avoiding any mounds or cracks in the soil.

2.2. Flight campaign and imagery quality parameters

The flight campaigns were performed with a high-resolution digital RGB camera mounted in two different types of UASs across the growing cycles: the fixed-wings ($Plat_{FW}$) eBee (SenseFly Ltd., Cheseaux-Lausanne, Switzerland) employed in Y17 and Y18, and multi-rotors ($Plat_{MR}$) AscTec Falcon 8 (Ascending Technologies, Krailling, Germany) in Y17 and Matrice 100 (DJI, Nanshan, Shenzhen, China) in Y18.

The flights were planned at the time of PHground phenotyping for assessing trials according to the predominant GS of interest in this study (E+40, B, A+7, and M). Table 2 summarizes the number of flights and main specifications for each GS assessment at the time of PHaerial estimation, including the number of flights. The ground control points (GCPs) or post-processed kinematic (PPK; see below for details) were used for georeferencing corrections. A set of black and white squared GCPs were uniformly distributed over the entire field area in all trials. These GCPs, distributed for each panel according to Table 2, were surveyed with a Global Navigation Satellite System (GNSS) receiver using a real-time kinematic (RTK) correction (Trimble R4 GNSS system, Trimble, Sunnyvale, CA, USA). Additionally, 11 checkpoints (CP) surveyed using RTK correction, were placed across the site during the crop cycle Y18 for georeferenced accuracy assessment of the orthomosaics.

Table 2. Crop phenology information across the measurements presented as days after emergence (DAE), the predominant development crop stage expressed by Zadoks growth scale, corresponding phenological stage and identification nomenclature in this investigation, as well as the number of flights for each platform using ground control points (GCPs) or post-processed kinematic (PPK) corrections for fixed-wing ($Plat_{FW}$) and multi-rotor ($Plat_{MR}$) platforms

Trial	GS	Ident. Stage[1]	Zadoks Scale[2]	DAE[3]	N of flights ($Plat_{FW}$)		N of flights ($Plat_{MR}$)
					PPK	GCP	Only GCP
HiBAP-I Y17 (30/Nov/2016[a])	Stem elongation	E+40	37-39	40	1	0	1
	Flowering	A+7	61-65	73-87	5	0	5
	Maturity	M	91-92	100	1	0	1
HiBAP-II Y18 (03/Dec/2017)	Stem elongation	E+40	37-39	40	1	0	0
	Booting	B	41-47	55-72	6	0	7
	Flowering	A+7	61-69	76-98	6	2	8
	Maturity	M	91-92	105-118	2	0	3
HiBAP-I Y18	Stem elongation	E+40	37-39	40	1	0	1

(18/Dec/2017)	Booting	B	41-45	55-69	2	1	3
	Flowering	A+7	61-69	74-91	1	1	2
	Maturity	M	91-92	106-111	1	0	1

[1] Specific identification of the growth stage (GS) estimated/predominant: at 40 days after emergence (E+40), booting (B), seven days after anthesis (A+7) and at physiological maturity (M); [2] The decimal (or Zadoks Scale) growth stage code estimated according to the genetic variability; [3] Days after emergence represented by 50% of the plants with the first leaf through coleoptile (GS10); [a] Emergence date in each crop season.

The flights of the $Plat_{FW}$ followed the technical recommendations in Loladze et al. (2019) and are described in Table 3. The flight plan was designed for north/south and east/west flights, to achieve both a lateral and longitudinal overlap of 80%. The flights covered an area larger than the experiment to cover the entire experimental field and obtain accurate orthomosaics. High-accuracy corrections of the geolocation data measured with the $Plat_{FW}$ global navigation satellite system (GNSS) were calculated in the post-processing using as a reference the position of a fixed base station and the PPK correction while imagery geotagging (Benassi et al., 2017; Forlani et al., 2018).

Table 3. Parameters of flight specifications details for fixed-wind ($Plat_{FW}$) and multi-rotor ($Plat_{MR}$) platforms.

	HiBAP-I Y17		**HiBAP-II Y18**		**HiBAP-I Y18**	
	$Plat_{FW}$	$Plat_{MR}$	$Plat_{FW}$	$Plat_{MR}$	$Plat_{FW}$	$Plat_{MR}$
Sensor	Canon Power Shot 110camera of 16.2 MegaPixels	Sony NEX 5	SODA	Zen Muse X5	SODA	Zen Muse X5
Resolution (Image Pixels)	4608 × 3456	4592 × 3056	5472 × 3648	4608 × 3456	5472 × 3648	4608 × 3456
GSD[1] Resolution (cm/Pixel)	1.7	0.7	1.7	0.7	1.7	0.7
GCPs[2] numbers for internal processing	7	7	9	9	7	7
Flight altitude	65	30	85	30	85	30

[1]GSD: ground sampling distance; [2]GCPs: ground control points used for internal processing for $Plat_{MR}$ and $Plat_{FW}$ without PPK corrections.

The flight plans for both multi-rotor platforms were designed to achieve lateral and longitudinal overlaps of 80%, flying north/south. The flight operations of these multi-rotor UASs are shown in Table 3, and further details can be checked in Tattaris et al. (2016) for the

AscTec Falcon 8, and in Horton and Ranganathan. (2018) for the Matrice 100. The flight plans of both types of platforms, $Plat_{FW}$ and $Plat_{MR}$, were designed to acquire images with different ground sampling distances (GSD in Table 3).

2.3. Three-dimensional crop reconstruction and Plant height accuracy assessment

The aerial data collected by both types of platforms were geotagged for orthomosaic processing using Pix4D Mapper® software (v4.4.10; Pix4D, Lausanne, Switzerland). Images were imported into Pix4D software, and for those $Plat_{FW}$ flights that did not use PPK corrections, as well as the flight campaign using $Plat_{MR}$, GCPs were manually located to improve the accuracy of the three-dimensional (3D) point cloud georeferencing (Figure 1, A). The digital terrain model (DTM, i.e., the topography of the site without any plant) was generated for each trial from images collected by a single flight of each UAS platform prior to the vegetation emergence. The digital surface model (DSM; i.e., the topography of the site accounting for the plants) was obtained along with vegetation development at each GS.

The DSM and DTM rasters were computed following the workflow recommended by Pix4D for high-resolution RGB imagery (Pix4D, 2019b). This workflow uses a structure from motion (SfM) algorithm (Ullman, 1979; Snavely et al., 2008) to obtain a 3D point cloud. The point cloud was later meshed via an algorithm based on Delauney triangulation (Matthew et al., 2009; Susanto et al., 2016) computed on multiple image scales with noise filtering and a 'sharp' surface smoothing filter. Afterward, the DTM was subtracted from the DSM to estimate the crop surface model (CSM, i.e., the height of individual plot surfaces) using R software version 3.6.1. (Development Core Team, 2018). The PHaerial scripts and processes used to perform the image analyses and trait extract are available at https://github.com/volpatoo/HTP-via-drone-imagery/tree/master/UAV-HTP_PlantHeight. Figure 1 represents the major steps of the data acquisition and processing, software, packages, and remote sensing tools used in this workflow.

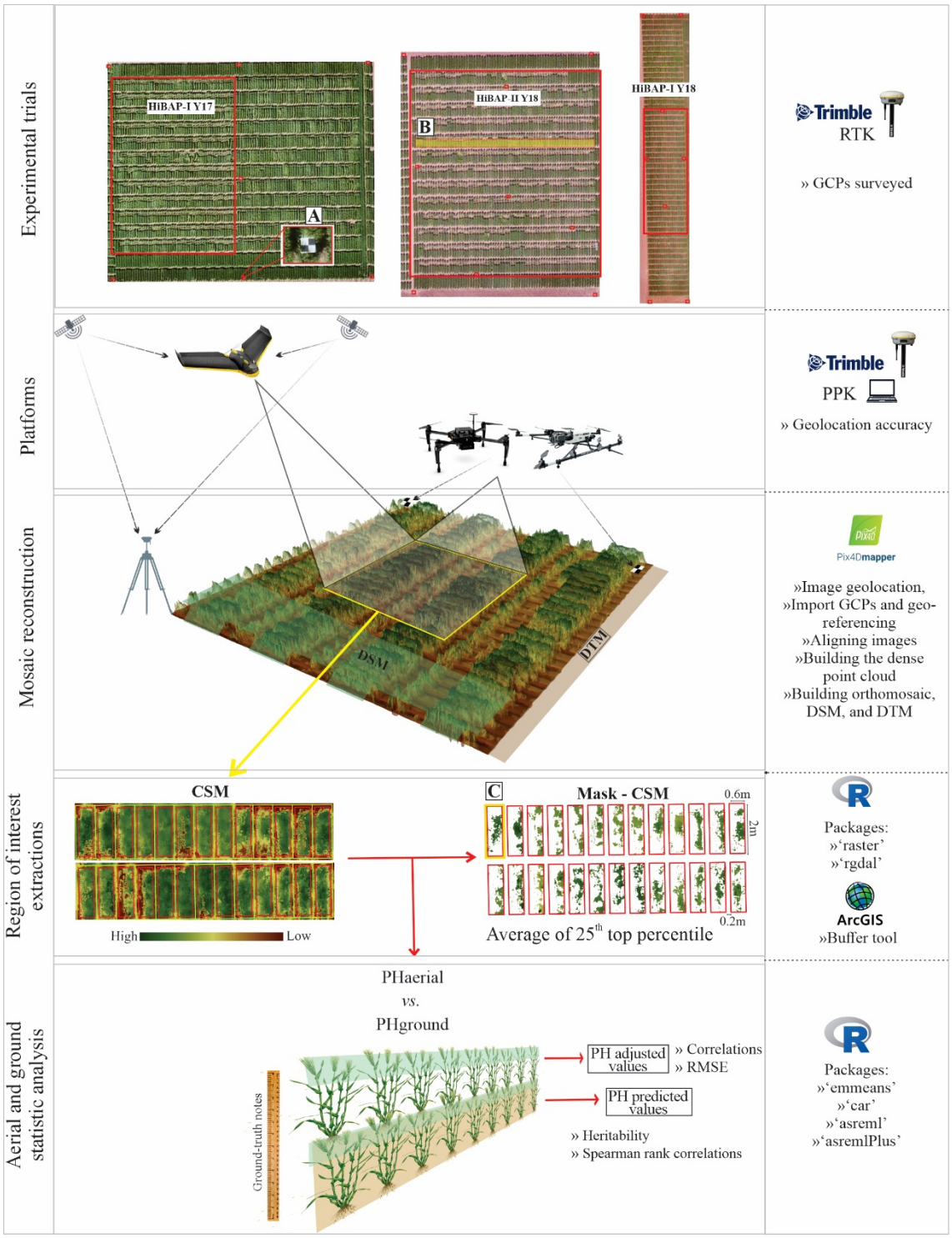


Figure 1. Phenotyping workflow for estimation of plant height (PH) using $Plat_{FW}$ and $Plat_{MR}$ (UASs) platforms (PHaerial) used in north-west Mexico during 2017 and 2018 growing cycles. PHground: ground-truth measurements; DTM: digital terrain model; DSM: digital surface model; CSM: crop surface model. The letters A, B, and C are GCPs design, locality of the profile selected, and cropped area with the mask of CSM as the top 25th percentile pixels value, respectively.

Before computing the CSM for all flight campaigns, a preliminary test to define the best parametrization of the Pix4D workflow was conducted. Different settings in Pix4D were combined and tested to obtain the best high-density point clouds and DSM (for details see Supplementary Table S2). The test results (not shown) were compared based on the accuracy of PHaerial against the PHground for each platform. For this exercise, the data from the Y17 growing cycle at E+40, A+7 and M GS, was used. The best performing processing scheme parameters were used to generate the 3D point clouds of all the flight campaigns (Supplementary Table S2)

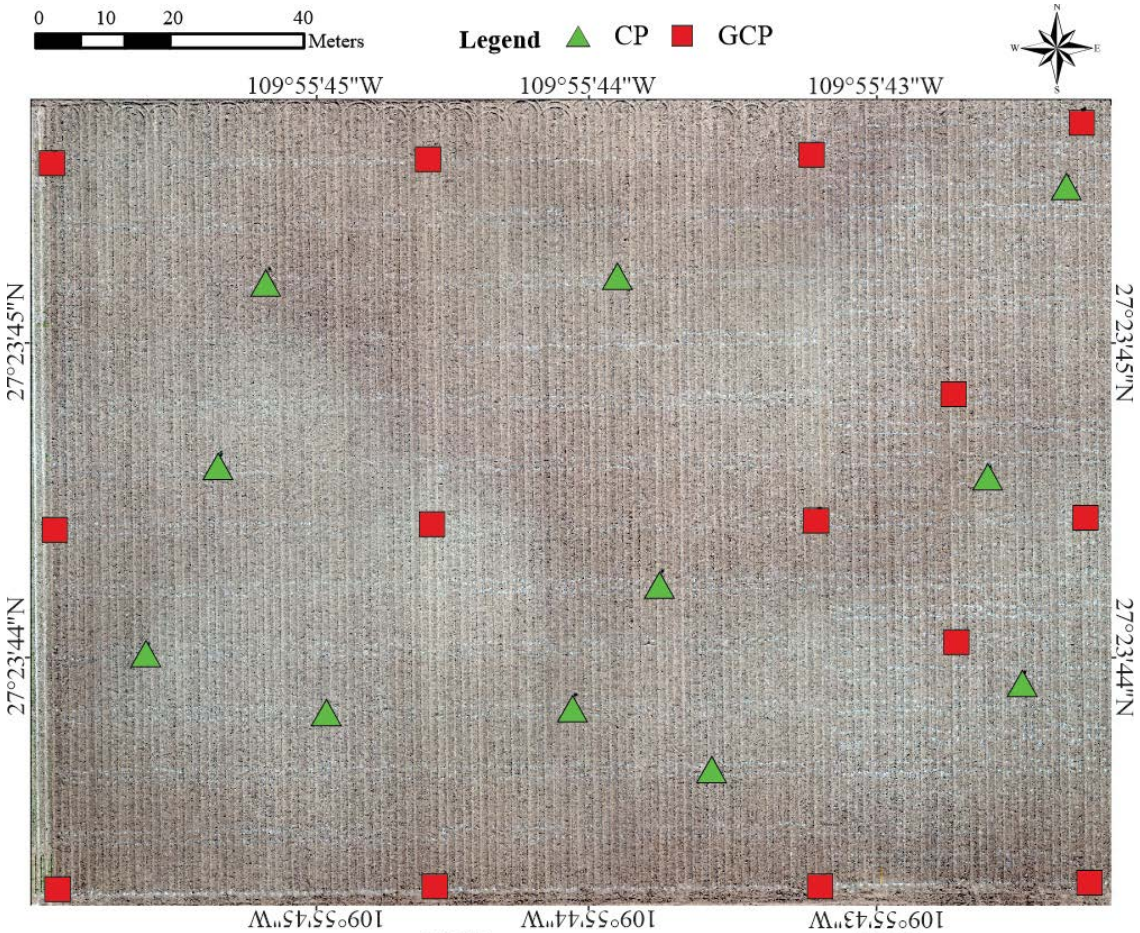
The PH from the CSM raster was assessed using ArcGIS (version 10.6, Esri Inc., Redlands, USA). The buffer tool was used to create regions of interest (ROI_{PH}) to extract PH values from each plot (Figure 1, C). Plant breeding trials usually consist of small plots within 0.5 to 1 m of each other in the interests of trial uniformity. Under these conditions the canopies of adjacent plots can interfere with one another by shading, lodging, or wind load. The small plots can easily cause noise in the PH estimation, especially after flowering. To ensure the extraction of pure pixel values (i.e., pixels containing only information from the plot of interest), the ROI_{PH} were built considering a buffer zone of 0.1 m from the plot edges, and were aligned at the center of the two-bed rows. ROI_{PH} were exported as polygons into a shapefile for the data extraction.

The data extraction per ROI_{PH} was achieved by overlapping the CSM and the shapefile containing the ROI_{PH} using the R packages 'raster' and 'rgdal'. Average PHaerial was calculated for each plot using pixel values greater than the 75th percentile for that plot. We tested different criteria for selecting pixels within the ROI_{PH} but this proved to be the optimum indicator for PHaerial based on comparison with the PHground values.

2.4. Accuracy assessment of orthomosaics georeferencing

The automation of data extraction per plot requires a high accuracy in the orthomosaics and DTM georeferencing. To ensure that, a preliminary study was performed using two techniques: GCPs and PPK correction. PPK correction was used to obtain accurately geotagged $Plat_{FW}$ imagery. The table in Figure 2 shows the comparison between the absolute accuracy of longitude and latitude coordinates estimated by the two methods. Absolute accuracy validation was performed using the CPs (which were not used in the bundle adjustment process) by comparing the coordinates of the CPs in the CSM with the reference values measured in-site by RTK GNSS (i.e., delta-XY, being X latitude and Y longitude). The delta-XY was calculated

for both platforms using the set of 11 CPs placed in the field during the crop cycle Y18. Additionally, the root mean square error (RMSE) of the differences between X and Y coordinates, the mean values and the standard deviations (SD) were computed. These parameters showed that the PPK achieved similar results than those obtained with GCPs for horizontal XY coordinates (RMSE ~ 1 cm and SD < 3/62 cm; Figure 2). The average accuracy measured as SD on the CPs coordinates was in agreement with the accepted limits mentioned by Vautherin et al. (2016): 1–2 times the GSD in X and Y directions either to GCP or PPK corrections.



	GCP		PPK	
	X(cm)	Y(cm)	X(cm)	Y(cm)
Mean[1]	-0.8184	0.6810	-0.3274	1.7980
SD	3.0794	2.5758	3.2832	3.6209
RMSE	0.9285	0.7766	0.9899	1.0917

Figure 2. The mean delta-XY from all geocoordinates obtained from 11 checkpoints (CP) versus the geocoordinates obtained from the orthomosaic (CSM) resulted from 14 GCP and PPK corrections. Standard deviation (SD) and Root Mean Square Error (RMSE) for X and Y coordinates. The base image corresponds to the bare soil flight using $Plat_{FW}$ in the HiBAP-II trial during November 2018. [1]Mean difference between measured coordinates to GCP and PPK

The accuracy for altitude estimations (i.e., Z-axis) was also evaluated by contrasting calculated and ground-truth GCPs height values using one flight in each breeding cycle for $Plat_{FW}$ and $Plat_{MR}$ (Fig 3). The height accuracy measured on the GCPs was acceptable in all flight dates, with the $Plat_{MR}$ showing slightly better results (RMSE = 1.77 to 1.85; and SD = 1.63 to 1.76) than $Plat_{FW}$ (RMSE = 2.81 to 3.84; and SD = 1.62 to 2.88). The R^2 was greater than 0.95 for all cases. The accuracy measured as SD also attended the criterion adopted by Vautherin et al. (2016): 2–3 times the GSD in the Z direction for both platforms. Overall, the accuracy obtained in the CSM using PPK and GCP approaches reached similar results.

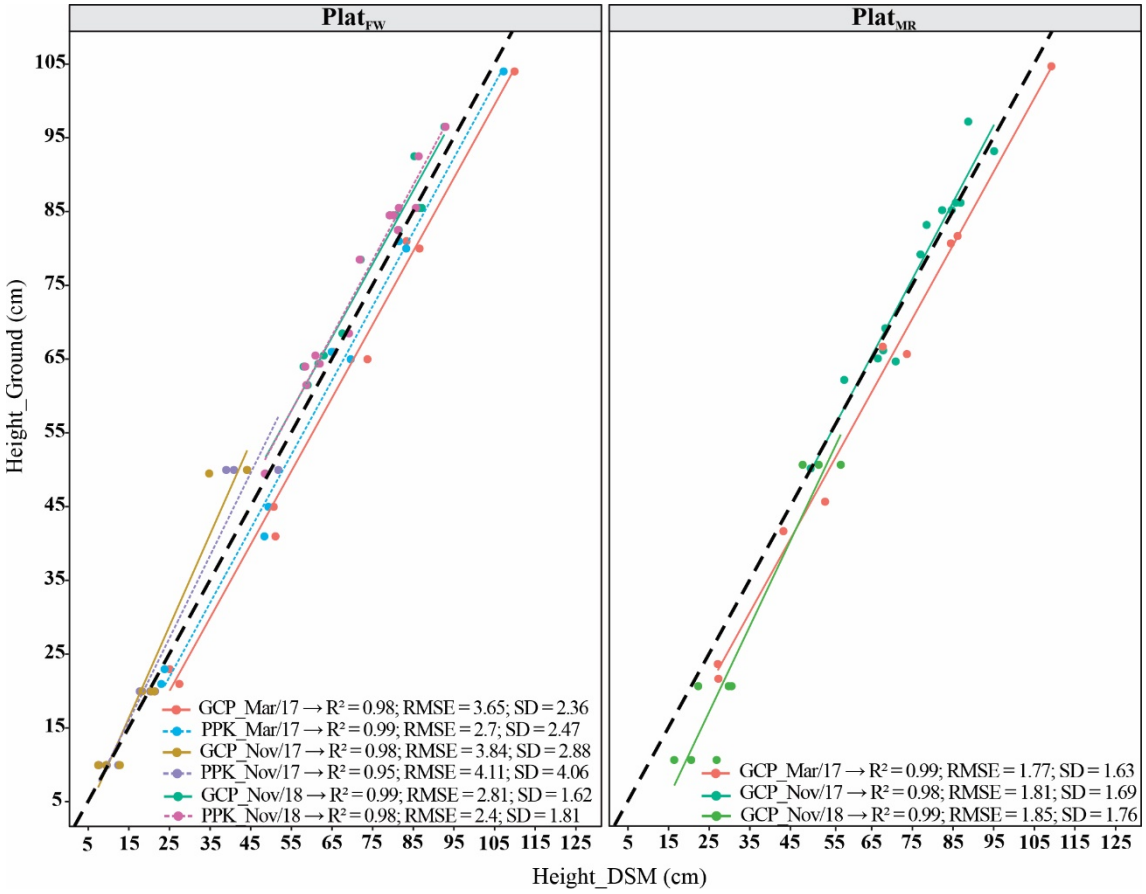


Figure 3. Coefficient of determination (R^2), root mean square deviation (RMSE), and standard deviation (SD) of the errors, between ground control points (GCPs) height estimate from the digital surface model (DSM), and ground-truth measurements for three different dates of flights using PPK for fixed-wind platform ($Plat_{FW}$) and only GCP corrections for $Plat_{FW}$ and multi-rotor ($Plat_{MR}$). The dark black dashed diagonals represent 1:1 line, and the colored solid (GCP) and dashed (PPK) lines represent the regression lines.

2.5. Statistical Models and Genetic selection evaluation

Since few flights were performed at each GS, a linear model was first built to fit a single PH value per plot. In this model, the PHaerial from each platform at each GS was used as a dependent variable against plot and number of flights as explanatory variables. The adjusted means per plot for PHaerial and PHground values were then used to calculate best linear unbiased estimates (BLUEs) within each crop cycle, using the following model:

$$y_{ijk} = \mu + g_i + r_j + b_{k(j)} + \varepsilon_{ijk}$$

where y_{ijk} is the trait value for genotype *i*, replicate *j*, and block *k*; μ is the overall mean; g_i is the fixed effect for genotype *i*; r_j is the random effect for replicate *j*, which are assumed to be independently and identically distributed according to a normal distribution with mean zero and variance σ_r^2; that is, $r_j \sim iid\ N(0, \sigma_r^2)$; $b_{k(j)} \sim iid\ N(0, \sigma_b^2)$ is the random effect for block *k* within replicated *j;* and $\varepsilon_{ijk} \sim iid\ N(0, \sigma_\varepsilon^2)$ is the residual effect.

For each growing cycle, Pearson's correlations, R^2, and RMSE for PHs between PHground and PHaerial were calculated using the BLUEs derived from the above model at each GS after removing the outliers. Outliers were flagged using studentized residual from PHground values, and the significance of its correlation with PHaerial was by the Bonferroni test at P<0.01 (Fox and Weisberg, 2019). The identified outliers were removed from both PHground and PHaerial to perform the analysis. Finally, the $RMSE_{dev}$ was computed to measure the deviation between the estimated values (PHaerial) and the measured values (PHground) across GS in each trial, according to Zhou et al. (2020).

The validation of the prediction model was done using best linear unbiased predictions (BLUPs) and heritability for PHground and PHaerial. The genotypic variance components (σ_g^2 and σ_ε^2) were derived by the fitted model above for both PHground and PHaerial to calculate the broad-sense heritability (H_g^2, sometimes termed "repeatability") with the genotype g_i treated as a random effect in which $g_i \sim iid\ N(0, \sigma_g^2)$. Thus, H_g^2 quantifying the repeatability of the plant height trait estimation was computed as the ratio between the genotypic to the total variances (Holland et al., 2002). The significance (Ripley, 2019) of the Spearman rank correlation coefficient (ρ) (Spearman, 1904) was calculated using the BLUPs from both UAS-platforms against PHground for assessing the accuracy of genotypic rank selection.

Additionally, fifty coincident genotypes measured in Y17 and Y18 (HiBAP-I) cycles were used considering $g_i \sim iid\ N(0, A\sigma_g^2)$ to fit the model, and obtain the narrow-sense heritability (h_a^2) for both UAS-platforms to also assess the accuracy under a GxE interaction model design. In this case, $\boldsymbol{A}$ is the associated additive relationship matrix, and by including the genotype × year interaction effect (t_{ge}) also as random with $t_{ge} \sim N(0, \sigma_{ge}^2)$. The data collected during booting in HiBAP-I Y18 was removed from the statistic-genetic model for G×E interaction in order to match better the GSs and calculate the BLUPs. R^2 represents the accuracy of predicted values from the correlations between the PHground and PHaerial. The standard errors (SE) of the heritability parameters in both validation models were obtained through mixed model output (Wolak, 2018).

The R software was used to run the statistical analyses, including linear models (Gilmour et al., 2015), multiple comparison procedures (Lenth, 2016), mixed and prediction models (Brien, 2018), and testing of model terms;; (Fox et al., 2019). The coefficients of parentage for the pedigree relationship matrices (*A*) were estimated as twice the coefficient of parentage using the "Browse" application within the International Crop Information System software package (McLaren et al., 2000).

3. Results

3.1. Descriptive Statistics across growth stages

PHground values were similar across the crop cycles for each GS (Figure 4). The heterogeneity within each trial remained relatively stable at B, A+7, and M (SD = 4.13 to 4.97 in HiBAP-I Y17, SD= 6.02 to 7.04 in HiBAP-I Y18, and SD= 4.62 to 5.65 in HiBAP-II Y18). The median value and SD for ground-truth PH measured at E+40 showed some discrepancies across cycles and trials, possibly attributed to the different genotypes used in each HiBAP panel, and due to the year effect and differences in emergence date.

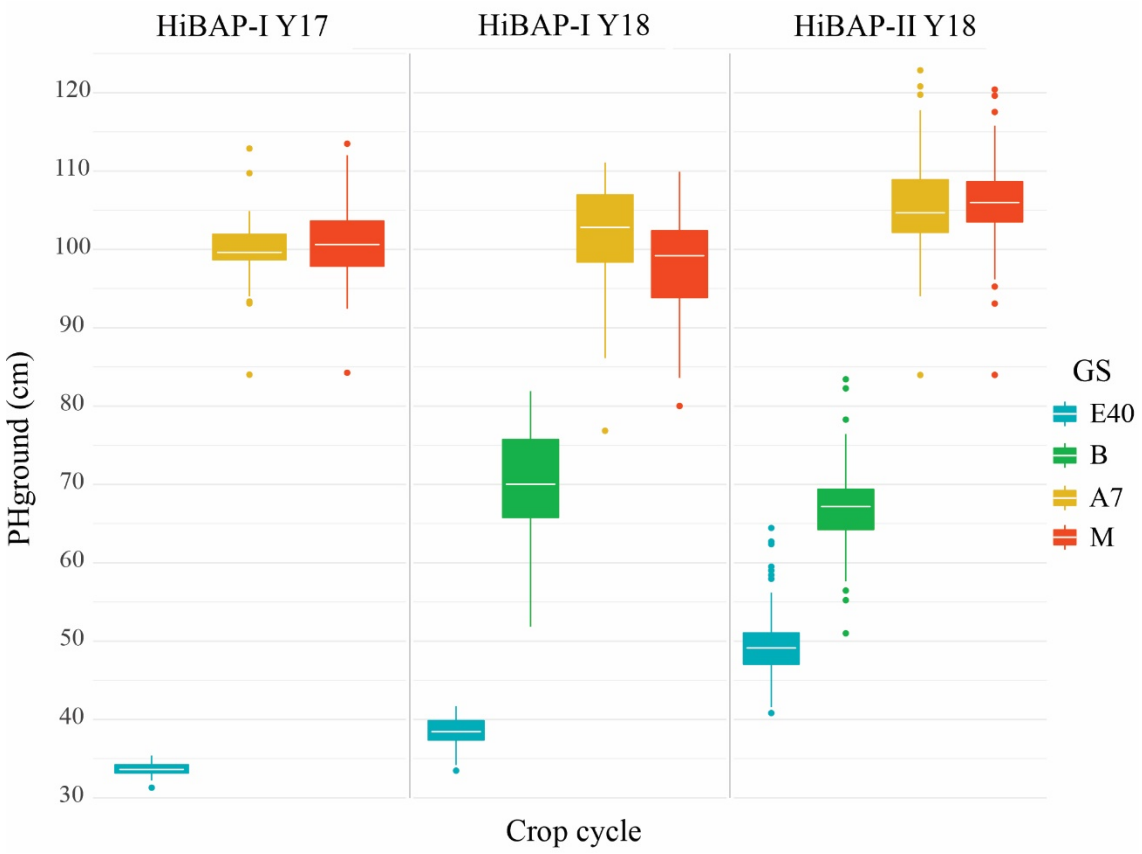


Figure 4. Boxplot of BLUEs for wheat plant height (PH, cm) in each of the three observed crop cycles at the following growth stages (GS): 40 days after emergence (E+40), booting (B), seven days after flowering (A+7) and at physiological maturity (M).

3.2. UAS plant height assessment and validation

The PHaerial estimates were, in general, similar to PHground values. This matching can be visualized in Figure 5, where transects of PHground and PHaerial data from HiBAP-II are compared (refer to Fig 1-B for the location of this transect within the HiBAP-II trial). Considerable mismatching between PHground and PHaerial values were detected at booting (B), whereas the best agreements were observed during maturity (M).

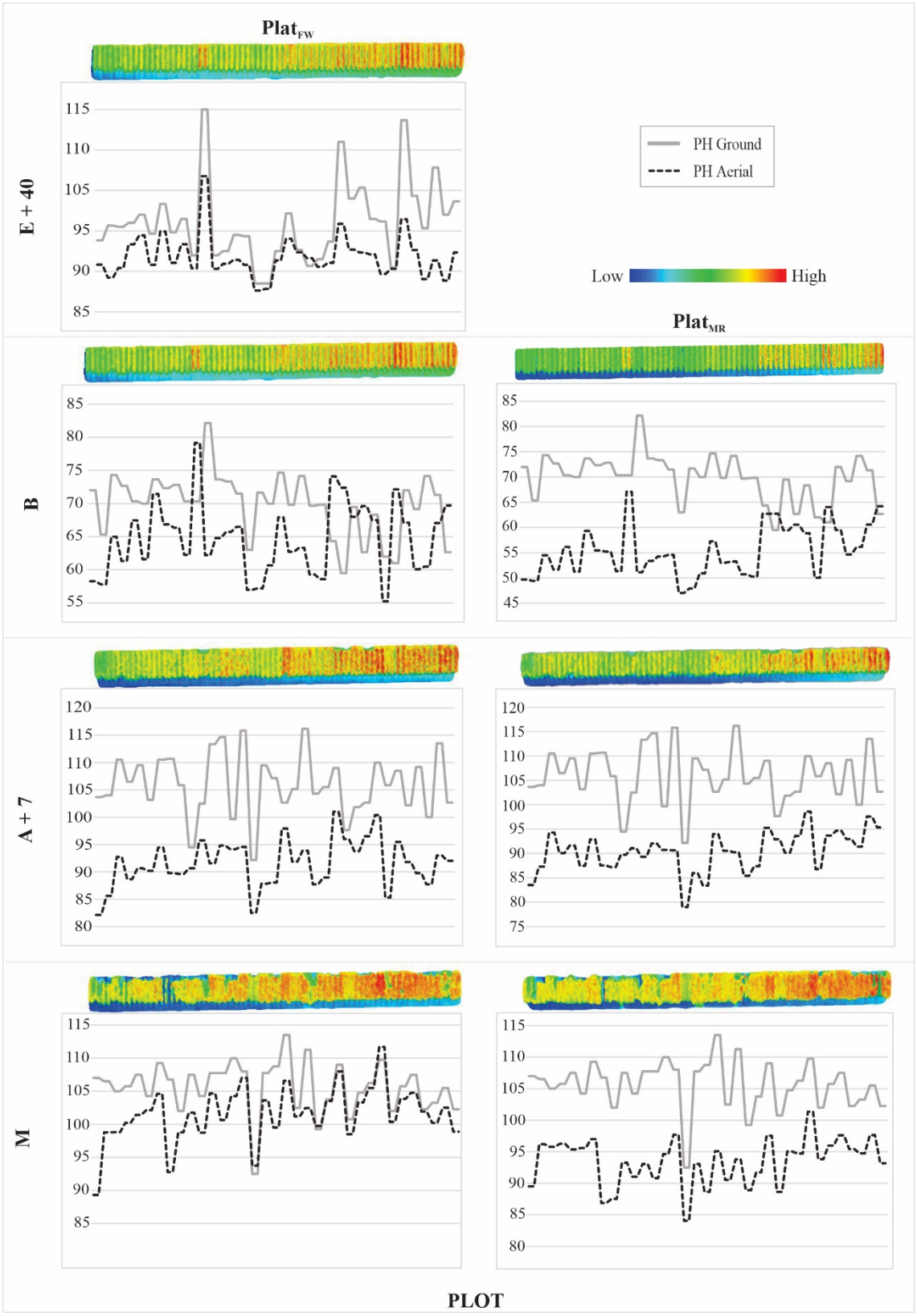


Figure 5. Profile of 70 plots from HiBAP-II for $Plat_{FW}$ and $Plat_{MR}$ at the followinggrowth stages (GS): 40 days after emergence (E+40), booting (B), seven days after flowering (A+7), and at physiological maturity (M). Plant height (PH, cm) via ground-truth (ground) and unmanned aerial system (UAS) data (aerial) are represented in the solid and dotted lines, respectively, matching with low (blueish) and high (reddish) color scale to PH. The 2D plots

profiles image were generated using the textured mesh feature via the densified point cloud of Pix4D processing.

The agreements between PHground and PHaerial data were further confirmed by the strong correlations observed during most of the growing cycle for both platforms (Figure 6). The coefficient of determination at the different GS ranged from non-significant to $R^2 = 0.88$, and from non-significant to $R^2 = 0.81$, for $Plat_{FW}$ and $Plat_{MR}$, respectively. The measurements at maturity showed the highest and most consistent correlations across the different trials and platforms, with R^2 values ranging between 0.7 and 0.88. The lowest correlations were obtained at booting, observing even non-significance in HiBAP-II Y18 for both $Plat_{FW}$ and $Plat_{MR}$ (*p-value* ≥ 0.78). The $Plat_{FW}$ platform performed better than $Plat_{MR}$ in HiBAP-I, except at E+40 during Y17, where $Plat_{MR}$ showed a coefficient of determination of 0.66 against 0.41 in $Plat_{FW}$. Conversely, $Plat_{MR}$ performed slightly better than $Plat_{FW}$ in HiBAP-II (Fig. 6), particularly at A+7 ($R^2 = 0.47$ in $Plat_{MR}$ vs $R^2 = 0.37$ in $Plat_{FW}$) and M ($R^2 = 0.74$ in $Plat_{MR}$ vs $R^2 = 0.7$ in $Plat_{FW}$). Overall, the RMSE of the predicted model for individual GS did not exceeded 4.02 cm. However, the $RMSE_{dev}$ obtained across GS for each platform in HiBAP-I Y17 and HiBAP-I Y18 were around 15 cm ($RMSE_{dev}$ = 15.06 and 14.95 cm in HiBAP-I Y17; $RMSE_{dev}$ = 14.44 and 15.42 in HiBAP-I Y 18). The best performance for $RMSE_{dev}$ was in HiBAP-II Y18 in both platforms. Nevertheless, the $Plat_{FW}$ provided better results than $Plat_{MR}$ ($RMSE_{dev}$ = 8.19 for $Plat_{FW}$ *vs.* 12.14 = $Plat_{MR}$).

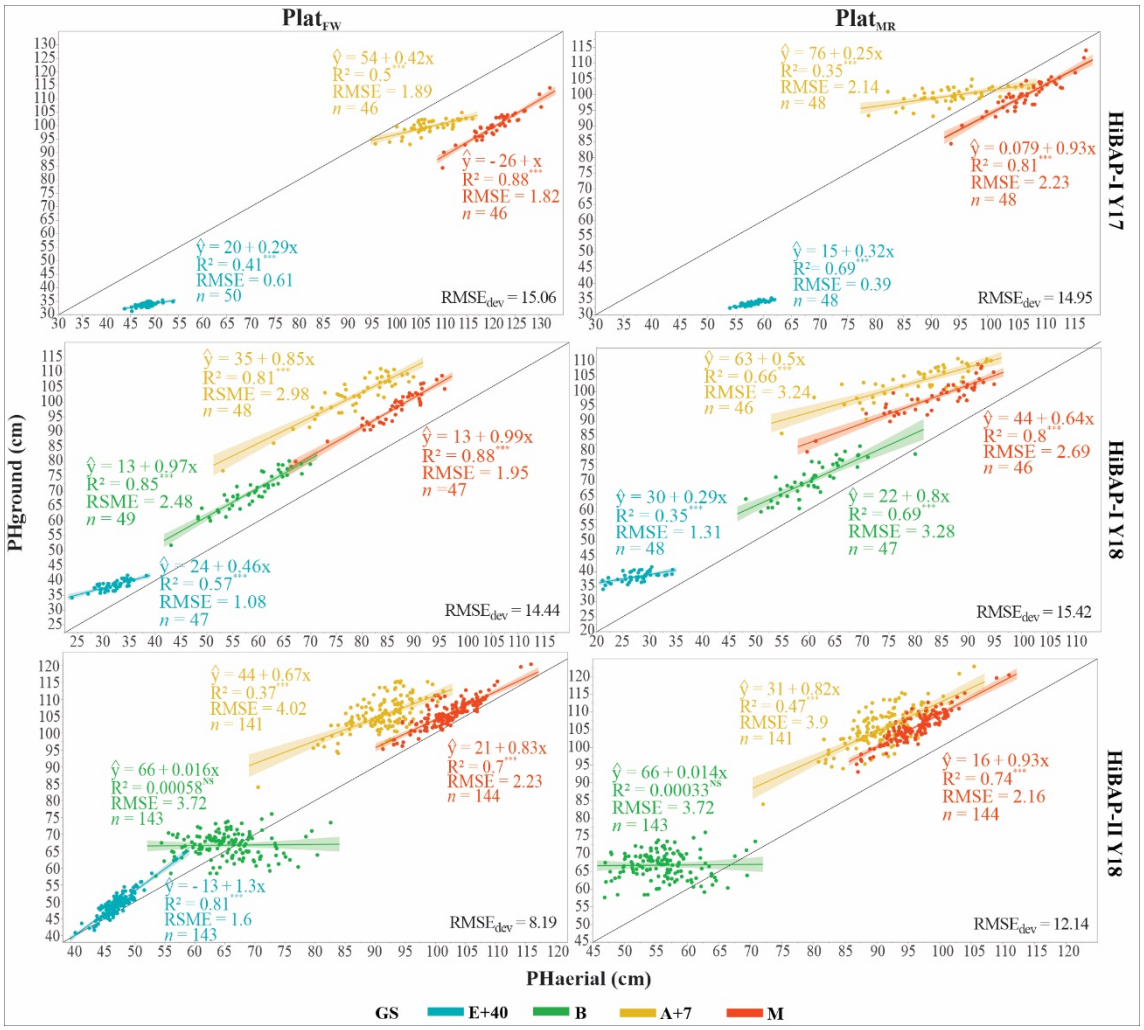


Figure 6. Linear relationship between plant height (PH, cm) estimated from the unmanned aerial system (UAS-based) data (PHaerial) and that ground-truth measured manually (PHground), as well as RMSE, and the number (n) of genotypes considered at four growth stages (GS): 40 days after emergence (E+40), booting (B), seven days after flowering (A+7), and at physiological maturity (M) for $Plat_{FW}$ and $Plat_{MR}$ considering all locations in this study. Black solid line shows the 1:1 lines; light shadow color represents a 99% confidence interval. The $RMSE_{dev}$ in the bottom right represents the deviation between the PHaerial and the PHground across GS. *. **, *** indicate p-value of the coefficient of determination (R^2), with * $P < 0.05$, ** $P < 0.01$ and *** $P < 0.001$; NS: non-significative value.

3.3. HTP for Genotypic Prediction of Plant Height from Wheat Breeding Trials

The evaluation strategy using H_g^2 shows strong potential for the PHaerial implementations in a wheat breeding program, reaching similar or higher values than those from PHground for each GS and across locations (Figure 7).

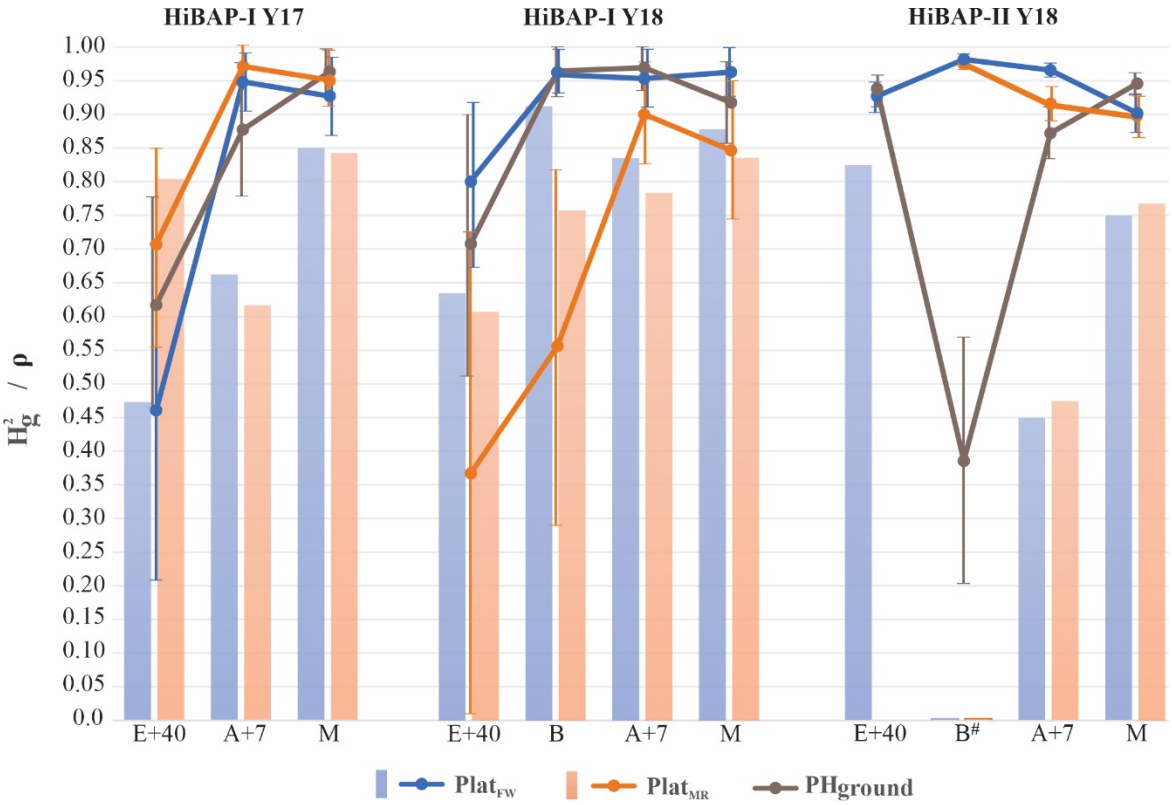


Figure 7. The solid color lines represent the broad-sense heritability (H_g^2) across crop cycles for $Plat_{FW}$, $Plat_{MR}$ and PHground in the growth stages: 40 days after emergence (E+40), booting (B), seven days after flowering (A+7), and at physiological maturity (M) and its 95% confidence interval based on standard errors. The color bars show the Spearman Rank Correlation (ρ) from the predicted values between PHground and PHaerial. All ρ significant at $P < 0.0001$ except to B (non-significant) in HiBAP-II Y18

The $Plat_{MR}$ performed better H_g^2 estimations than $Plat_{FW}$ and PHground for all GS at HiBAP-I Y17, except for M from PHground, ranging from 0.71 to 0.97 for $Plat_{MR}$ *vs.* 0.46 to 0.93 for $Plat_{FW}$, and *vs.* 0.62 to 0.96 for PHground. On the other hand, the $Plat_{FW}$ obtained greater H_g^2 values than $Plat_{MR}$ and PHground at HiBAP-I Y18 in all GS analyzed, except for A+7 from

PHground (H_g^2 = 0.80, 0.96, 0.95, and 0.90 for $Plat_{FW}$ vs H_g^2 = 0.37, 0.56, 0.90, and 0.85 for $Plat_{MR}$, and vs H_g^2 = 0.71, 0.96, 0.97, and 0.92 for PHground, for E+40, B, A+7 and M, respectively). On HiBAP-II Y18, both platforms obtained similar results, but significantly better than PHground at B according to the confidence interval (CI), as well as $Plat_{FW}$ at A+7 better than $Plat_{MR}$ and PHground. Furthermore, $Plat_{FW}$ and PHground in HiBAP-I Y18 were significantly better than $Plat_{MR}$ at B.

Overall, the H_g^2 responses were in agreement with the results from the correlations (R^2) between PHground and PHaerial data, which for HiBAP-I trials the H_g^2 degraded at E+40, and it increases at later GS. Additionally, the average H_g^2 values estimated across GS for each trial reveal that UASs platforms were better than PHground (H_g^2 = 0.78, 0.92 and 0.94 for $Plat_{FW}$, and H_g^2 = 0.88, 0.67 and 0.93 for $Plat_{MR}$ vs. H_g^2 = 0.82, 0.89 and 0.78 for PHground, within HiBAP-I Y17, HiBAP-I Y18, and HiBAP-II Y18 trials, respectively).

The Spearman rank correlations (ρ) between predicted values for PHaerial and PHground was significant (P<0.0001) at all GS in all trials except at booting in HiBAP-II Y18 where it was not significant. The highest ρ for HiBAP-I Y17 was observed at M for both platforms, and at HiBAP-I Y18, except at B using the $Plat_{FW}$ (ρ = 0.91). Moreover, the greatest ρ in HiBAP-II Y18 was achieved at E+40 via $Plat_{FW}$ (ρ = 0.83). Lower, but still significant correlations using both platforms were observed at A+7 in HiBAP-II Y18 (ρ = 0.45 for $Plat_{FW}$, and ρ = 0.46 for $Plat_{FW}$) (Figure 7).

When the genotype-environment interaction (G×E) effects were considered to predict the genotypic PH values, the narrow-sense heritability (h_a^2) for $Plat_{FW}$ was greater than for $Plat_{MR}$ for all GS analyzed (h_a^2 = 0.29, 0.65 and 0.62 in $Plat_{FW}$, vs h_a^2 = 0.06, 0.42 and 0.41 in $Plat_{MR}$, for E+40, A+7, and M growth stages, respectively). However, the h_a^2 from PHground were higher than PHaerial at A+7 (h_a^2 = 0.71) and M (h_a^2 = 0.71). The accuracy (or R^2) remains constant across GS ranging from 0.75 to 0.96 in $Plat_{FW}$ vs 0.64 to 0.92 in $Plat_{MR}$, but with lower values at E+40 for both UAS-platforms (Figure 8).

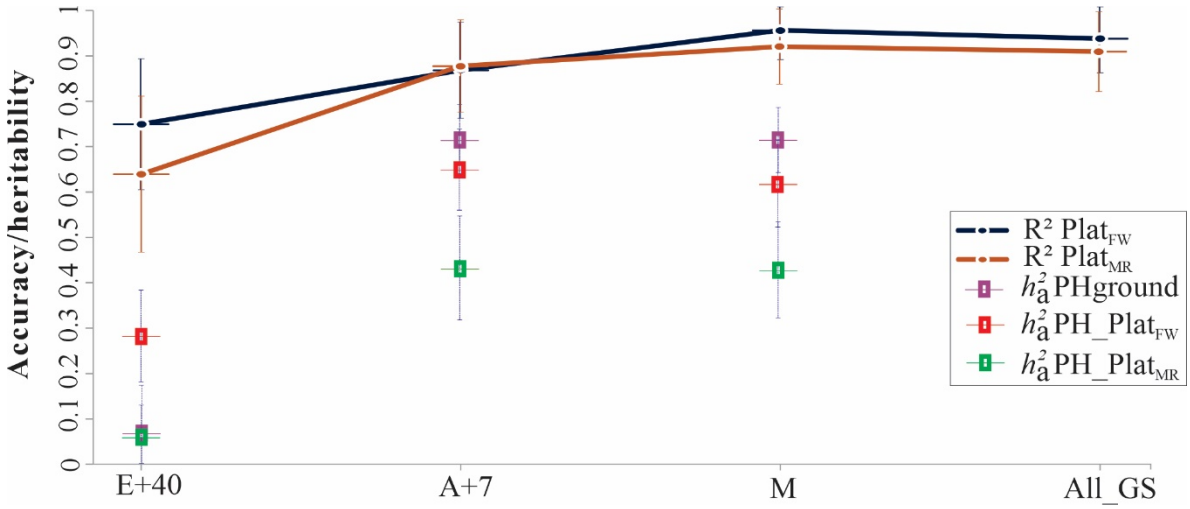


Figure 8. Narrow-sense heritability (h_a^2) and coefficient of determination (R^2) from genotype-by-environment interaction (G×E) model using HiBAP-I genotypes measured in 2016-2017 and 2017-2018 growing cycles via $Plat_{FW}$ and $Plat_{MR}$, and across growth stages (GS): 40 days after emergence (E+40), seven days after flowering (A+7), and at physiological maturity (M). The solid color lines represent the accuracy (R^2) for predict values between PHground and PHaerial, plus the confidence interval (CI) by the error bar. The colored squared represents the h_a^2 and the error bars indicate standard error. All growth stages (All_GS) were also considered to confirm the accuracy of genetic correlations. The cross lines give the upper, and lower bounds of the 95% CI of the correlations computed using $\hat{\rho} \pm 1.96\sqrt{(1-\hat{\rho})/(n-2)}$, where $\hat{\rho}$ is the estimated correlation, and n is the number of records used to compute the correlation.

4. Discussion

The present study aimed to prove the applicability of aerial photogrammetry (i.e., using UASs) to estimate PH in the wheat breeding context. Throughout the growing season, aerial HTP approaches were conducted on two different wheat panels (HiBAP-I and II) with two UAS-platform classes (multi-rotor and fixed-wing) equipped with RGB cameras using PPK or GCPs corrections. Our findings showed that for most of the growing stages, the UAS-based data (PHaerial) could be used for reliable estimations of PH and that genotype selection based on this data was equivalent to that obtained by manual ground measurements (PHground) (R^2 = 0.35 – 0.88). We attribute the good results obtained for a large number of experimental wheat plots (100 for HiBAP-I and 450 for HiBAP-II), partly to the strategy used in the pixel PH values extraction within each plot. The selection of pixels from the top 25th percentile within each plot ROI of the imagery was intended to increase the proportion of observations from the upper canopy in further analyses (Figure 1, C). This strategy showed a reliable method also used by other authors (Table 1), however contrasting in selected percentiles (e.g., 99th or 99.5th).

4.1. Assessing the data quality of the UAS based plant height

The differences observed in performance between the two aerial platforms suggest that UAS imagery parameters such as GSD, altitude, and point cloud density may influence their capability for accurate estimations of PH. Several authors have discussed the importance of GSD to create high-quality orthomosaics via orthorectification to obtain the DSM from the dense image matching as an additional source of data to enhance the PH model accuracy (Madec et al., 2017; Lu et al., 2019; Wilke et al., 2019). This orthomosaic generation method confers more accurate 3D points due to extraction of common characteristic points (keypoints) in different images and removes perspective distortion from the images using the DSM (Pix4D, 2019a). Our results indicate the $Plat_{FW}$ produced more accurate PH despite having a higher GSD (GSD = 2 cm/pixel), and a lower point cloud density (see Supplementary Table 1) compared to $Plat_{MR}$.

Even though $Plat_{FW}$ produced the best correlations and RMSE results overall in this study, the choice between the two classes of evaluated platforms depends on the processing pipeline used, plant breeding target, and several other technical factors such as area extension, pilot expertise, total flight time, and intended GSD (Petrie, 2013; DroneDeploy, 2017; Puri et al., 2017; Park et al., 2019; Raeva et al., 2019). The high-resolution images recorded from the $Plat_{MR}$ captured a higher variability within the vegetation, probably making the CSM reconstruction more susceptible to slight plant movements and changes in illumination conditions within the canopy. Hassan et al. (2019a) also discussed few sources of error for aerial estimation of PH in wheat crops from inefficient image pre- and post-processing due to suboptimal flying altitude, inaccurate DTM construction, and height extraction strategy from images. Chu et al. (2017) demonstrated that under unfavorable weather conditions, the quality of a dense point could affect the 3D-pixel constructions, and Song and Wang (2019) provided a suitable method that might overtake the PH bias eliminating noise from UAS-based 3D point cloud datasets for wheat fields. In this regard, biased PH data in wheat can occur due to wind conditions when using high spatial resolutions images similar to those recorded from $Plat_{MR}$ (GSD = 0.7 cm/pixel), as well as indicated by Willkomm et al. (2016) study.

The environmental conditions during data acquisition can also lead to noisy point clouds that hinder PH estimations. This factor would result in a lower accuracy of 3D point during the

orthorectification processing, affecting the point cloud densification step (Tirado et al. 2019). Indeed, our results for HiBAP-I Y17 show evidence that the PHaerial accuracy increased as the wind speed decreased, in contrast with the solar radiation that influenced the correlations (but not significative differences across GS according to the CI) in HiBAP-I Y18 (Supplementary Figure 1). Other environmental factors that can potentially have an impact on the PHaerial are shadows (Jin et al., 2017; Brocks and Bareth, 2018), radiometric calibration (Mafanya et al., 2018), brightness levels (López-Granados et al., 2019), and cloudy weather (Niedzielski and Jurecka, 2018).

The UAS/PPK results showed a high agreement with those obtained from CPs and GCPs that could be an affordable method to potentially increasing the image georeference accuracy by reducing the human interference (i.e., survey GCPs; import, and manual mark GCPs into the software) (Figures 2 and 3). Jin et al. (2017) concluded that using high-resolution images and low altitude flights via $Plat_{MR}$ induce potential errors for the photogrammetric techniques using GCPs. However, the calibration of the camera from the bundle adjustment in the use of UAS imagery for HTP approaches also depends on a proper georeferencing of the images, which is improved by a proper distribution of a sufficient number of GCPs across the field (Harwin and Lucieer, 2012; Sanz-Ablanedo et al., 2018) and a high overlapping between images (Madec et al., 2017) as elaborated in this study. The RTK/PPK has been used on-board by some authors (Benassi et al., 2017; Forlani et al., 2018; Sanz-Ablanedo et al., 2018; Padró et al., 2019) to correct the location of drone mapping imagery to improve the accuracy of GNSS data or geotags as they are captured either during flights (RTK) or after flight (PPK). However, based on their conclusions, it is unclear whether direct georeferencing using RTK/PPK will supersede GCPs to become the standard referencing technique for UAS imagery. The challenges of surveying the GCPs and keeping them statically placed throughout the life of the crop, makes the PPK method cost-effective. It is also time-consuming to check all GCPs on the photogrammetry software to computes the keypoints on the images. To avoid this issue in a practical field situation, we recommend the use of a UAS/PPK system as implemented for $Plat_{FW}$, since the results were comparable to those orthomosaics georeferenced with only GCPs.

Finally, the settings used in the software for orthomosaics and the DSM generation would also impact the accuracy of PH estimation using UASs. As discussed by Holman et al. (2016), the software settings should be carefully selected and accurately reported for futures improvements in

UAS based SfM methods. The current study achieved reliable outputs after testing different combinations of the settings in Pix4D. Additionally, we provide a user-fridely script to perform the pixels value extractions using an open-source software (R software), and its insights to automate the CSM acquisition using a loop throughput DSM images.

4.2. Accuracy and phenotypic variations in UAS-based plant height

The strong correlations and CI observed between PHground and PHaerial values, indicate that this approach can be used for growth rate analysis and wheat selection in a breeding pipeline. The UAS data processing chain used in this study demonstrated itself to be quick, cost-effective, and accurate for the target trait. Moreover, our results showed similar accuracy levels than previous studies (Table 1) for PH estimation at individual GS, in which the correlations were higher at late GS, as obtained in some cases (Hassan et al., 2019a). Nevertheless, it is essential to note that in some studies the authors report correlations across stages along the growth cycle (Madec et al., 2017; Yue et al., 2017; Yuan et al., 2018; Harkel et al., 2019; Lu et al., 2019), masking the actual assessment power of PHaerial by stretching the response variable (i.e., ground PHs).

High-throughput estimations of PH have also achieved reliable results using other platforms in several crops (Malambo et al., 2018; Wang et al., 2018; Harkel et al., 2019; Reynolds et al., 2020). In particular, ground-based light detection and ranging (LiDAR) has been reported to give more accurate PH estimations than UAS photogrammetry in wheat (Holman et al., 2016; Iqbal et al., 2017; Madec et al., 2017; Jimenez-Berni et al., 2018; Deery et al., 2020). However, the implementation of such a platform is still not high versatile and can be expensive (Nex and Remondino, 2014; Guo et al., 2018) in contrast with high-resolution RGB cameras. Lately, cutting-edge technologies in remote sensing have triggered the rapid development of affordable high-performance sensors (i.e., RGB, multispectral and hyperspectral cameras) and UASs with higher autonomy and payload capacity, increasing the possibilities for field phenotyping applications (Sankaran et al. 2015). Our UAS approach is comparatively cheap and more efficient for phenotyping large and multi-location trials, targeting quick data acquisition and reducing computer resources; concepts that are supported by the literature via SfM approaches (Yeyin et al., 2016; Wang et al., 2018; Reynolds et al., 2019).

Despite the high correlations and acceptable RSME of the fitted model, in early stages (i.e., E+40 in Figure 6), the limited range of variation of PH can limit the correlations as demonstrated in HiBAP-I trials (Figure 4) and confirmed by Madec et al. (2017). In some cases, our results either underestimated (HiBAP-I Y18 and HiBAP-II Y18, Figure 6) or overestimate (HiBAP-I Y17, Figure 6) the absolute values of PH compared to the PHground. These results can also be observed by $RMSE_{dev}$, which corroborates with errors in the literature to wheat PH estimations (Madec et al., 2017; Jimenez-Berni et al., 2018; Yuan et al., 2018; Hassan et al., 2019a). The $RMSE_{dev}$ assessed by the agreement between PHground and PHaerial also was used in this study as a measure of accuracy. Bias of crop heights by UAS platforms could be due to the lack of ability of SfM to reconstruct the top of the canopy accurately (Madec et al. 2017), influence of neighboring plants (Khanna et al., 2015; Watanabe et al., 2017), and the inaccurate DTM strategy to pixel value extraction (Hu et al., 2018).. Willkomm et al. (2016) also described that plant movement during image acquisition, and the lack of a adjusted protocol to in-field measurements led to PH underestimations. In this context, using fixed-ground targets at different heights proved to be an essential validation step in the current data processing workflow for PH estimations. The very accurate height estimations of the ground targets achieved using $Plat_{MR}$ and $Plat_{FW}$ (R^2> 0.95 and RMSE < 4.11 cm; Figure 3), show the real potential of this method, and suggest that differences in plot-level estimations of PH between the PHaerial and PHground may be partly, but not solely, related to inaccuracies in the manual scouting over very extensive field trials, as well as due to wind movement.

In our study, the lack of correlations in HiBAP-II at booting highlighted some issues on PHground, and this factor is easily detectable when drawing a transect to compare trend lines (Figure 5). Accurate phenotyping is fundamental either for the calibration or validation of a novel HTP approaches (Araus and Cairns, 2014), as well as reported in studies for high-throughput genotyping (Ma et al., 2014). Reynolds et al. (2019) discuss the cost-benefit for phenotyping, showing that UAS-based photogrammetry is relatively affordable when flights operate under favourable conditions, i.e., with no rain, sunny days, and light to moderate wind-speed. However, during the flights performed at B in HiBAP-II Y18, the weather conditions for wind speed and solar radiation were reasonable compared with the other GS in the same trial. The weak correlations in this case could be attributed to heterogeneity within and between plots (Rosyara et al., 2019) and

the canopy architecture issues to detecting the booting during the vegetative stage (Torres and Pietragalla, 2012) in experimental wheat plots.

The accuracy of the DTM is paramount for accurate estimations of PH (Bendig et al., 2014; Iqbal et al., 2017), especially in highly dense canopies such as those observed at A+7 and M (Yang et al., 2019) DTM can be generated along with the presence of vegetation, but generally at late GS, the bare soil is barely exposed close to the ROI during flights missions to in-field pixel capture. However, the DTM can also be obtained from measurements over the bare soil before the vegetation grows. The precision of the DTM estimated from the vegetation DSM depends on the number and distribution of bare ground patches across the experimental sites, and the capacity to detect them. Furthermore, the detection of bare soil pixels is usually done by image classification methods, which can be affected by rugged relief (Hassan et al., 2019a). Despite these limitations, some authors prefer the estimation of DTM from vegetation DSM, arguing advantages in terms of processing time (Zhang et al., 2018; Hassan et al., 2019a). In our study, we use as a baseline a DTM generated from bare soil images acquired before plant emergence. The advantage of this approach is that it does not rely on image classification algorithms, whose performance depends on the scenario (e.g., plot dimensions, crop development and growth stages of interest, etc.). A drawback of using such DTM is that it does not account for changes in the topography generated by the irrigation and personnel walking in the field, which can also account for weak PH estimations.

4.3. UAS-based plant height as a reliable trait for wheat phenotyping

The satisfactory correlations (R^2) between PHground and PHaerial observed in this study indicated the applicability of the current UAS-based workflow. However, plant breeders often calculate the heritability as a measure of the precision of trials and/or to compute the response to selection (Piepho and Möhring, 2007; Schmidt et al., 2019). In this regard, the heritability (H_g^2 and h_a^2) was used to confirm the UAS-based approach's ability to inferthe predicted genetic values. Different genotype responses in various environments by GS were also investigated(Piepho and Williams, 2006). Additionally, the genotypic correlation was used to compare the similarity between PHground and PHaerial rankings using the predicted values.

The highest H_g^2 value estimated, on average across GS, for PHaerial as compared to PHground could indicate better phenotype measurement performance. In this case, the ranking of the best genotypes according to the breeding program propose can be reached using PHaerial predictions. This finding was also confirmed by the Spearman rank correlation (Figure 7). Therefore, our workflow to phenotyping PH combined with reliable H_g^2 can be an affordable and efficient method to let breeders be more accurate in genotype selection criteria. Other studies also have supported higher heritability (or repeatability, in some cases) to PH in later GS (Novoselovic et al., 2004; Hassan et al., 2019a; Deery et al., 2020). However, some issues may appear in the temporal image when the target traits depend on the geometric structure, as described by (Madec et al., 2017), who observed poor H_g^2 at the end of the growth cycle due to plant lodging. These circumstances were not evident in our study. However, the weather conditions discussed in the previous session, has the potential to bias the results, as pointed out by PH correlations in this study.

Medium to low levels of h_a^2 observed in the GS suggest a meaningful environmental influence, indicating that the G×E interactions affect the PH predictions (Figure 8). These results were supported by the random effect significance (Wilks, 1938) of G×E interaction at most of the GS analyzed, except in E+40 for $Plat_{MR}$ and PHground (data not shown). The non-significance, in these cases, can be attributed in part due to the limited range of variability for PH. Furthermore, lower h_a^2 are expected when compared with H_g^2 due to pedigree information, in which the h_a^2 uses the proportion of genetic variation due to additive genetic effects only (Bernardo, 2020). The results confirmed that PH is a critical trait responsible for G×E interaction as expected to be in quantitative traits (Tian et al., 2017; Tshikunde et al., 2019). The G×E interaction can indicate better the cultivar performance and environments that facilitate genotype evaluation (Mohammadi et al., 2015). The high correlations between predicted values for PHaerial and PHground across and within GS indicate that each platform measured similar underlying genetic trait. This means PHaerial can reliably predict genotypic values and rank genotypes as reliable as PHground.

5. Conclusions

The present study implemented and validated an efficient and scalable approach to acquire PHs measurements under extensive wheat breeding trials at different growth stages. The remote sensing techniques applied in this study allowed the estimation of PHs using high-resolution RGB imagery

recorded from two UAS platforms and processed through a semi-automatic pipeline. The results for all trials in two growing cycles prove that the present workflow was able to estimate PHs from UAS-platforms comparable in accuracy to those measured by ground-truth measurements. Our findings also indicated that using PHaerial for genotype selection could be cost-effective to predict PHs values by temporal data of drone imagery from multiple environments, mainly in late GS. Due to the reliable results achieved by $Plat_{FW}$ to compute PH, it is acceptable to indicate that lower density point cloud does not confer PH noise or underestimation in comparison to $Plat_{MR}$. The accuracy was responsive to image quality (i.e., GSD, weather conditions, etc.,.) and the settings in the processing steps of the surface model generation. A proper georeferencing of the orthomosaic is an essential step for the data extraction, and the UAS-PPK approach demonstrated to be a suitable method to replace the laborious conventional methods using GCPs.

As evidenced by wheat PH studies in Table 1, different approaches for pixel extractions can be made choosing different thresholds for capturing the genotype variability within and among experimental plots. In this study, the reliable results via increasing temporal resolution throughout the GSs in wheat breeding trials also was endorsed by the authors in Table 1 under PH estimation ability by continuous capture of cost-effective phenotypic data in multiple environments. Finally, this study demonstrates that it is feasible to process reliably high-volume of field-based phenotypic data using UAS-based imagery.

6. Acknowledgments

This work was supported by the CGIAR Research Program on Wheat (www.wheat.org), by the International Wheat Yield Partnership (IWYP) and by the Sustainable Modernization of Traditional Agriculture (MasAgro) - an initiative from the Secretariat of Agriculture and Rural Development (SADER) and CIMMYT. We are very thankful to CIMMYT's Biometrics and Statistics Unit (BSU), for their support on statistical programming; for the whole Crop Nutrition and Wheat Physiology teams for their assistance with field measurements, flight campaign and image processing. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

7. References

Aisawi, K. A. B., Reynolds, M. P., Singh, R. P., and Foulkes, M. J. (2015). The physiological basis of the genetic progress in yield potential of CIMMYT spring wheat cultivars from 1966 to 2009. *Crop Sci.* 55, 1749–1764. doi:10.2135/cropsci2014.09.0601.

Araus, J., and Cairns, J. (2014). Field high-throughput phenotyping: the new crop breeding frontier. Trends Plant Sci. 19, 52–61.

Awika, J. M. (2011). Major Cereal Grains Production and Use around the World. Available at: https://pubs.acs.org/sharingguidelines.

Barmeier, G., Mistele, B., and Schmidhalter, U. (2016). Referencing laser and ultrasonic height measurements of barleycultivars by using a herbometre as standard. Crop Pasture Sci. 67, 1215. doi:10.1071/cp16238.

Benassi, F., Dall'Asta, E., Diotri, F., Forlani, G., Cella, U. M. di, Roncella, R., et al. (2017). Testing accuracy and repeatability of UAV blocks oriented with gnss-supported aerial triangulation. Remote Sens. 9. doi:10.3390/rs9020172.

Bendig, J., Bolten, A., Bennertz, S., Broscheit, J., Eichfuss, S., and Bareth, G. (2014). Estimating biomass of barley using crop surface models (CSMs) derived from UAV-based RGB imaging. Remote Sens. 6, 10395–10412. doi:10.3390/rs61110395.

Bernardo, R. (2010). Breeding for Quantitative Traits in Plants. 2nd Editio. University of Minnesota–Twin Cities: Stemma Press.

Bernardo, R. (2020). Reinventing quantitative genetics for plant breeding: something old, something new, something borrowed, something BLUE. Heredity (Edinb). doi:10.1038/s41437-020-0312-1.

Brien, C. (2018). asremlPlus: Augments the Use of "ASReml-R" and "ASReml4-R" in Fitting Mixed Models. R Packag V. Available at: https://cran.r-project.org/web/packages/asremlPlus/asremlPlus.pdf.

Brocks, S., and Bareth, G. (2018). Estimating barley biomass with crop surface models from oblique RGB imagery. Remote Sens. 10. doi:10.3390/rs10020268.

Chapman, S., Merz, T., Chan, A., Jackway, P., Hrabar, S., Dreccer, M., et al. (2014). Pheno-Copter: A Low-Altitude, Autonomous Remote-Sensing Robotic Helicopter for High-Throughput Field-Based Phenotyping. Agronomy 4, 279–301. doi:10.3390/agronomy4020279.

Chu, T., Starek, M. J., Brewer, M. J., Murray, S. C., and Pruter, L. S. (2017). Assessing lodging severity over an experimental maize (Zea mays L.) field using UAS images. Remote Sens. 9, 1–24. doi:10.3390/rs9090923.

Cobb, J. N., DeClerck, G., Greenberg, A., Clark, R., and McCouch, S. (2013). Next-generation phenotyping: Requirements and strategies for enhancing our understanding of genotype-

phenotype relationships and its relevance to crop improvement. Theor. Appl. Genet. 126, 867–887. doi:10.1007/s00122-013-2066-0.

Castro, A. I., Jiménez-Brenes, F. M., Torres-Sánchez, J., Peña, J. M., Borra-Serrano, I., and López-Granados, F. (2018). 3-D characterization of vineyards using a novel UAV imagery-based OBIA procedure for precision viticulture applications. Remote Sens. 10. doi:10.3390/rs10040584.

De Souza, C. H. W., Lamparelli, R. A. C., Rocha, J. V., and Magalhães, P. S. G. (2017). Height estimation of sugarcane using an unmanned aerial system (UAS) based on structure from motion (SfM) point clouds. Int. J. Remote Sens. 38, 2218–2230. doi:10.1080/01431161.2017.1285082.

Deery, D. M., Rebetzke, G. J., Jimenez-berni, J. A., Condon, A. G., Smith, D. J., Bechaz, K. M., et al. (2020). Ground-Based LiDAR Improves Phenotypic Repeatability of Above-Ground Biomass and Crop Growth Rate in Wheat. Plant Phenomics 2020, 11. doi:https://doi.org/10.34133/2020/8329798.

DroneDeploy (2017). Choosing the Right Mapping Drone for Your Business Part I: Multi-Rotor vs. Fixed Wing Aircraft. Available at: https://blog.dronedeploy.com/choosing-the-right-mapping-drone-for-your-business-part-i-multi-rotor-vs-fixed-wing-aircraft-6ec2d02eff48 [Accessed July 25, 2019].

Forlani, G., Dall'Asta, E., Diotri, F., di Cella, U. M., Roncella, R., and Santise, M. (2018). Quality assessment of DSMs produced from UAV flights georeferenced with on-board RTK positioning. Remote Sens. 10. doi:10.3390/rs10020311.

Fox, J., and Weisberg, S. (2019). An R Companion to Applied Regression. Third. Thousand Oaks CA: Sage Available at: https://socialsciences.mcmaster.ca/jfox/Books/Companion/.

Fox, J., Weisberg, S., and Price, B. (2019). Package' car.' 608. Available at: https://r-forge.r-project.org/projects/car/, https://CRAN.R-project.org/package=car,%0Ahttps://socserv.socsci.mcmaster.ca/jfox/Books/Companion/index.html.

Gilmour, A. R., Gogel, B. J., and Welham, S. J. (2015). ASReml User Guide Functional Specification. doi:10.1186/1471-2458-7-114.

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

Guo, Q., Wu, F., Pang, S., Zhao, X., Chen, L., Liu, J., et al. (2018). Crop 3D—a LiDAR based platform for 3D high-throughput crop phenotyping. Sci. China Life Sci. 61, 328–339. doi:10.1007/s11427-017-9056-0.

Han, L., Yang, G., Yang, H., Xu, B., Li, Z., and Yang, X. (2018). Clustering Field-Based Maize Phenotyping of Plant-Height Growth and Canopy Spectral Dynamics Using a UAV Remote-Sensing Approach. Front. Plant Sci. 9. doi:10.3389/fpls.2018.01638.

Harkel, J. ten, Bartholomeus, H., and Kooistra, L. (2019). Biomass and Crop Height Estimation of Different Crops Using UAV-Based Lidar. Remote Sens. 14. doi:doi:10.3390/rs12010017.

Harwin, S., and Lucieer, A. (2012). Assessing the accuracy of georeferenced point clouds produced via multi-view stereopsis from Unmanned Aerial Vehicle (UAV) imagery. Remote Sens. 4, 1573–1599. doi:10.3390/rs4061573.

Hassan, M. A., Yang, M., Fu, L., Rasheed, A., Zheng, B., Xia, X., et al. (2019a). Accuracy assessment of plant height using an unmanned aerial vehicle for quantitative genomic analysis in bread wheat. Plant Methods 15. doi:10.1186/s13007-019-0419-7.

Hassan, M. A., Yang, M., Rasheed, A., Yang, G., Reynolds, M., Xia, X., et al. (2019b). A rapid monitoring of NDVI across the wheat growth cycle for grain yield prediction using a multispectral UAV platform. Plant Sci. 282, 95–103. doi:10.1016/j.plantsci.2018.10.022.

Hedden, P. (2003). The genes of the Green Revolution. Trends Genet. 19.

Hickey, L. T., N. Hafeez, A., Robinson, H., Jackson, S. A., Leal-Bertioli, S. C. M., Tester, M., et al. (2019). Breeding crops to feed 10 billion. Nat. Biotechnol. doi:10.1038/s41587-019-0152-9.

Hoffmann, A. A., and Merilä, J. (1999). Application of the load flow and random flow models for the analysis of power transmission networks. Reliab. Eng. Syst. Saf. 14, 102–109. doi:10.1016/j.ress.2012.02.005.

Holland, J. B., Nyquist, W. E., and Cervantes-Martínez, C. T. (2002). Estimating and Interpreting Heritability for Plant Breeding: An Update. , ed. J. Janick Plant Breeding Reviews doi:10.1002/9780470650202.ch2.

Holman, F. H., Riche, A. B., Michalski, A., Castle, M., Wooster, M. J., and Hawkesford, M. J. (2016). High throughput field phenotyping of wheat plant height and growth rate in field plot trials using UAV based remote sensing. Remote Sens. 8. doi:10.3390/rs8121031.

Horton, E., and Ranganathan, P. (2018). Development of a GPS spoofing apparatus to attack a DJI Matrice 100 Quadcopter. J. Glob. Position. Syst. 16. doi:10.1186/s41445-018-0018-3.

Hu, P., Chapman, S. C., Wang, X., Potgieter, A., Duan, T., Jordan, D., et al. (2018). Estimation of plant height using a high throughput phenotyping platform based on unmanned aerial vehicle and self-calibration: Example for sorghum breeding. Eur. J. Agron. 95, 24–32. doi:10.1016/j.eja.2018.02.004.

Iqbal, F., Lucieer, A., Barry, K., and Wells, R. (2017). Poppy crop height and capsule volume estimation from a single UAS flight. Remote Sens. 9, 24–27. doi:10.3390/rs9070647.

James, M. R., and Robson, S. (2014). Mitigating systematic error in topographic models derived from UAV and ground-based image networks. Earth Surf. Process. Landforms 39, 1413–1420. doi:10.1002/esp.3609.

Jimenez-Berni, J. A., Deery, D. M., Rozas-Larraondo, P., Condon, A. (Tony) G., Rebetzke, G. J., James, R. A., et al. (2018). High Throughput Determination of Plant Height, Ground Cover,

and Above-Ground Biomass in Wheat with LiDAR. Front. Plant Sci. 9. doi:10.3389/fpls.2018.00237.

Jin, X., Liu, S., Baret, F., Hemerlé, M., and Comar, A. (2017). Estimates of plant density of wheat crops at emergence from very low altitude UAV imagery. Remote Sens. Environ. 198, 105–114. doi:10.1016/j.rse.2017.06.007.

Khanna, R., Martin, M., Pfeifer, J., Liebisch, F., Walter, A., and Siegwart, R. (2015). Beyond Point Clouds - 3D Mapping and Field Parameter Measurements using UAVs. doi:10.1109/ETFA.2015.7301583.

Lenth, R. V. (2016). Least-Squares Means: The R Package lsmeans. J. Stat. Softw. 69. doi:10.18637/jss.v069.i01.

Li, J., Veeranampalayam-Sivakumar, A. N., Bhatta, M., Garst, N. D., Stoll, H., Stephen Baenziger, P., et al. (2019). Principal variable selection to explain grain yield variation in winter wheat from features extracted from UAV imagery. Plant Methods 15, 1–13. doi:10.1186/s13007-019-0508-7.

Loladze, A., Rodrigues, F. A., Toledo, F., San Vicente, F., Gérard, B., and Boddupalli, M. P. (2019). Application of Remote Sensing for Phenotyping Tar Spot Complex Resistance in Maize. Front. Plant Sci. 10. doi:10.3389/fpls.2019.00552.

López-Granados, F., Torres-Sánchez, J., Jiménez-Brenes, F. M., Arquero, O., Lovera, M., and De Castro, A. I. (2019). An efficient RGB-UAV-based platform for field almond tree phenotyping: 3-D architecture and flowering traits. Plant Methods 15, 1–16. doi:10.1186/s13007-019-0547-0.

Lu, N., Zhou, J., Han, Z., Li, D., Cao, Q., Yao, X., et al. (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 15. doi:10.1186/s13007-019-0402-3.

Ma, C., Zhang, H. H., and Wang, X. (2014). Machine learning for Big Data analytics in plants. Trends Plant Sci. 19, 798–808. doi:10.1016/j.tplants.2014.08.004.

Madec, S., Baret, F., de Solan, B., Thomas, S., Dutartre, D., Jezequel, S., et al. (2017). High-Throughput Phenotyping of Plant Height: Comparing Unmanned Aerial Vehicles and Ground LiDAR Estimates. Front. Plant Sci. 8. doi:10.3389/fpls.2017.02002.

Mafanya, M., Tsele, P., Botai, J. O., Manyama, P., Chirima, G. J., and Monate, T. (2018). Radiometric calibration framework for ultra-high-resolution UAV-derived orthomosaics for large-scale mapping of invasive alien plants in semi-arid woodlands: Harrisia pomanensis as a case study. Int. J. Remote Sens. 39, 5119–5140. doi:10.1080/01431161.2018.1490503.

Makanza, R., Zaman-Allah, M., Cairns, J. E., Magorokosho, C., Tarekegne, A., Olsen, M., et al. (2018). High-throughput phenotyping of canopy cover and senescence in maize field trials using aerial digital canopy imaging. Remote Sens. 10. doi:10.3390/rs10020330.

Malambo, L., Popescu, S. C., Murray, S. C., Putman, E., Pugh, N. A., Horne, D. W., et al. (2018). Multitemporal field-based plant height estimation using 3D point clouds generated from small unmanned aerial systems high-resolution imagery. Int. J. Appl. Earth Obs. Geoinf. 64, 31–42. doi:10.1016/j.jag.2017.08.014.

Matias, F. I., Caraza-harter, M. V, and Endelman, J. B. (2020). FIELDimageR : An R package to analyze orthomosaic images from agricultural field trials. 1–6. doi:10.1002/ppj2.20005.

Matthew, B., Michael, K., Randal, B., and Hugues, H. (2009). Parallel Poisson Surface Reconstruction. doi:10.1007/978-3-642-10520-3.

McLaren, C. G., Ramos, L., Lopez, C., and Eusebio, W. (2000). "Applications of the genealogy management system," in International Crop Information System. Technical Development Manual, eds. C. G. McLaren, J. W. White, and P. N. Fox (Mexico: CIMMYT and IRRI), 5.8–5.13. doi:10.1534/g3.118.200856.

Moeller, C., Sauerborn, J., de Voil, P., Manschadi, A. M., Pala, M., and Meinke, H. (2014). Assessing the sustainability of wheat-based cropping systems using simulation modelling: Sustainability = 42? Sustain. Sci. 9, 1–16. doi:10.1007/s11625-013-0228-2.

Mohammadi, R., Farshadfar, E., and Amri, A. (2015). Interpreting genotype × environment interactions for grain yield of rainfed durum wheat in Iran. Crop J. 3, 526–535. doi:10.1016/j.cj.2015.08.003.

Mohan, M., Silva, C. A., Klauberg, C., Jat, P., Catts, G., Cardil, A., et al. (2017). Individual tree detection from unmanned aerial vehicle (UAV) derived canopy height model in an open canopy mixed conifer forest. Forests 8. doi:10.3390/f8090340.

Molero, G., Joynson, R., Pinera-Chavez, F. J., Gardiner, L. J., Rivera-Amado, C., Hall, A., et al. (2019). Elucidating the genetic basis of biomass accumulation and radiation use efficiency in spring wheat and its role in yield potential. Plant Biotechnol. J. 17, 1276–1288. doi:10.1111/pbi.13052.

Morales, N., Mueller, L. A., Kaczmar, N. S., Robbins, K. R., Santantonio, N., and Gore, M. A. (2020). ImageBreed : Open-access plant breeding web – database for image-based phenotyping Field experiments. 1–7. doi:10.1002/ppj2.20004.

Moreira, F. F., Hearst, A. A., Cherkauer, K. A., and Rainey, K. M. (2019). Improving the efficiency of soybean breeding with high-throughput canopy phenotyping. Plant Methods 15, 1–9. doi:10.1186/s13007-019-0519-4.

Moreira, F. F., Oliveira, H. R., Volenec, J. J., Rainey, K. M., and Brito, L. F. (2020). Integrating High-Throughput Phenotyping and Statistical Genomic Methods to Genetically Improve Longitudinal Traits in Crops. 11, 1–18. doi:10.3389/fpls.2020.00681.

Nex, F., and Remondino, F. (2014). UAV for 3D mapping applications: A review. Appl. Geomatics 6, 1–15. doi:10.1007/s12518-013-0120-x.

Niedzielski, T., and Jurecka, M. (2018). Can Clouds Improve the Performance of Automated Human Detection in Aerial Images? Pure Appl. Geophys. 175, 3343–3355. doi:10.1007/s00024-018-1931-9.

Novoselovic, D., Baric, M., Drezner, G., Gunjaca, J., and Lalic, A. (2004). Quantitative inheritance of some wheat plant traits. Genet. Mol. Biol. 27, 92–98. doi:10.1590/S1415-47572004000100015.

Padró, J. C., Muñoz, F. J., Planas, J., and Pons, X. (2019). Comparison of four UAV georeferencing methods for environmental monitoring purposes focusing on the combined use with airborne and satellite remote sensing platforms. Int. J. Appl. Earth Obs. Geoinf. 75, 130–140. doi:10.1016/j.jag.2018.10.018.

Park, S., Lee, H., and Chon, J. (2019). Sustainable monitoring coverage of unmanned aerial vehicle photogrammetry according to wing type and image resolution. Environ. Pollut., 340–348. doi:10.1016/j.envpol.2018.08.050.

Peña-Bautista, R. J., Hernandez-Espinosa, N., Jones, J. M., Guzmán, C., and Braun, H. J. (2017). CIMMYT Series on Carbohydrates, Wheat, Grains, and Health: Wheat-Based Foods: Their Global and Regional Importance in the Food Supply, Nutrition, and Health , . Cereal Foods World 62, 231–249. doi:10.1094/cfw-62-5-0231.

Peña, J. M., Torres-Sánchez, J., Serrano-Pérez, A., de Castro, A. I., and López-Granados, F. (2015). Quantifying efficacy and limits of unmanned aerial vehicle (UAV) technology for weed seedling detection as affected by sensor resolution. Sensors (Switzerland) 15, 5609–5626. doi:10.3390/s150305609.

Petrie, G. (2013). Commercial Operation of Lghtweight UAVs for Aerial Imaging and Mapping:with Particular Reference to the U.K. GEOinformatics.

Piepho, H. P., and Möhring, J. (2007). Computing heritability and selection response from unbalanced plant breeding trials. Genetics 177, 1881–1888. doi:10.1534/genetics.107.074229.

Piepho, H. P., and Williams, E. R. (2006). A comparison of experimental designs for selection in breeding trials with nested treatment structure. Theor. Appl. Genet. 113, 1505–1513. doi:10.1007/s00122-006-0398-8.

Pix4D (2019a). Getting GCPs on the field or through other sources. Available at: https://support.pix4d.com/hc/en-us/articles/202557489-Step-1-Before-Starting-a-Project-4-Getting-GCPs-on-the-field-or-through-other-sources-optional-but-recommended [Accessed December 20, 2019].

Pix4D (2019b). How to improve the outputs of dense vegetation areas? – Support. Available at: https://support.pix4d.com/hc/en-us/articles/202560159-How-to-improve-the-outputs-of-dense-vegetation-areas [Accessed July 26, 2019].

Puri, V., Nayyar, A., and Raja, L. (2017). Agriculture drones: A modern breakthrough in precision agriculture. J. Stat. Manag. Syst. 20, 507–518. doi:10.1080/09720510.2017.1395171.

Radoglou-Grammatikis, P., Sarigiannidis, P., Lagkas, T., and Moscholios, I. (2020). A compilation of UAV applications for precision agriculture. Comput. Networks 172, 107148. doi:10.1016/j.comnet.2020.107148.

Raeva, P. L., Šedina, J., and Dlesk, A. (2019). Monitoring of crop fields using multispectral and thermal imagery from UAV. Eur. J. Remote Sens. 52, 192–201. doi:10.1080/22797254.2018.1527661.

Rebetzke, G. J., Jimenez-Berni, J., Fischer, R. A., Deery, D. M., and Smith, D. J. (2019). Review: High-throughput phenotyping to enhance the use of crop genetic resources. Plant Sci. 282, 40–48. doi:10.1016/j.plantsci.2018.06.017.

Reynolds, D., Baret, F., Welcker, C., Bostrom, A., Ball, J., Cellini, F., et al. (2019). What is cost-efficient phenotyping? Optimizing costs for different scenarios. Plant Sci. 282, 14–22. doi:10.1016/j.plantsci.2018.06.015.

Reynolds, M., Chapman, S., Crespo-Herrera, L., Molero, G., Mondal, S., Pequeno, D. N. L., et al. (2020a). Breeder friendly phenotyping. Plant Sci. 295, 110396. doi:10.1016/j.plantsci.2019.110396.

Reynolds, M. P., and Borlaug, N. E. (2006). Impacts of breeding on international collaborative wheat improvement. J. Agric. Sci. 144, 3–17. doi:10.1017/S0021859606005867.

Ripley, B. (2019). boot:Bootstrap Functions (. 1.3-23. Available at: https://cran.r-project.org/web/packages/boot/index.html.

Rodrigues, F. A., Blasch, G., Defourny, P., Ortiz-Monasterio, J. I., Schulthess, U., Zarco-Tejada, P. J., et al. (2018). Multi-temporal and spectral analysis of high-resolution hyperspectral airborne imagery for precision agriculture: Assessment of wheat grain yield and grain protein content. Remote Sens. 10. doi:10.3390/rs10060930.

Rosyara, U., Kishii, M., Payne, T., Sansaloni, C. P., Singh, R. P., Braun, H. J., et al. (2019). Genetic Contribution of Synthetic Hexaploid Wheat to CIMMYT's Spring Bread Wheat Breeding Germplasm. Sci. Rep. 9, 1–11. doi:10.1038/s41598-019-47936-5.

Sankaran, S., Khot, L. R., and Carter, A. H. (2015). Field-based crop phenotyping: Multispectral aerial imaging for evaluation of winter wheat emergence and spring stand. Comput. Electron. Agric. 118, 372–379. doi:10.1016/j.compag.2015.09.001.

Sanz-Ablanedo, E., Chandler, J. H., Rodríguez-Pérez, J. R., and Ordóñez, C. (2018). Accuracy of Unmanned Aerial Vehicle (UAV) and SfM photogrammetry survey as a function of the number and location of ground control points used. Remote Sens. 10. doi:10.3390/rs10101606.

Sayre, K. D., Rajaram, S., and Fischer, R. A. (1997). Yield potential progress in short bread wheats in northwest Mexico. Crop Sci. 37, 36–42. doi:10.2135/cropsci1997.0011183X003700010006x.

Schirrmann, M., Hamdorf, A., Garz, A., Ustyuzhanin, A., and Dammer, K. H. (2016). Estimating wheat biomass by combining image clustering with crop height. Comput. Electron. Agric. 121, 374–384. doi:10.1016/j.compag.2016.01.007.

Schmidt, P., Hartung, J., Bennewitz, J., and Hans-Peter, P. (2019). Heritability in plant breeding on a genotype-difference basis. Genetics 212, 991–1008. doi:10.1534/genetics.119.302134.

Shewry, P. R., Pellny, T. K., and Lovegrove, A. (2016). Is modern wheat bad for health? Nat. Plants 2. doi:10.1038/nplants.2016.97.

Singh, A., Ganapathysubramanian, B., Singh, A. K., and Sarkar, S. (2016). Machine Learning for High-Throughput Stress Phenotyping in Plants. Trends Plant Sci. 21, 110–124. doi:10.1016/j.tplants.2015.10.015.

Singh, D., Wang, X., Kumar, U., Gao, L., Noor, M., Imtiaz, M., et al. (2019). High-Throughput Phenotyping Enabled Genetic Dissection of Crop Lodging in Wheat. Front. Plant Sci. 10. doi:10.3389/fpls.2019.00394.

Snavely, N., Seitz, S. M., and Szeliski, R. (2008). Modeling the world from Internet photo collections. Int. J. Comput. Vis. 80, 189–210. doi:10.1007/s11263-007-0107-3.

Song, Q., Zhang, G., and Zhu, X. G. (2013). Optimal crop canopy architecture to maximise canopy photosynthetic CO 2 uptake under elevated CO2-A theoretical study using a mechanistic model of canopy photosynthesis. Funct. Plant Biol. 40, 109–124. doi:10.1071/FP12056.

Song, Y., and Wang, J. (2019). Winter Wheat Canopy Height Extraction from UAV-Based Point Cloud Data with a Moving Cuboid Filter. 10–14.

Spearman, C. (1904). The Proof and Measurement of Association between Two Things. Psychol. Bull. 1, 363–363. doi:10.1037/h0065390.

Susanto, F., de Souza, P., and He, J. (2016). Spatiotemporal Interpolation for Environmental Modelling. Sensors (Basel). 16, 1–20. doi:10.3390/s16081245.

Tattaris, M., Reynolds, M. P., and Chapman, S. C. (2016). A Direct Comparison of Remote Sensing Approaches for High-Throughput Phenotyping in Plant Breeding. Front. Plant Sci. 7, 1–9. doi:10.3389/fpls.2016.01131.

Tian, X., Wen, W., Xie, L., Fu, L., Xu, D., Fu, C., et al. (2017). Molecular mapping of reduced plant height gene Rht24 in bread wheat. Front. Plant Sci. 8, 1–9. doi:10.3389/fpls.2017.01379.

Tirado, S. B., Hirsch, C. N., and Springer, N. M. (2019). UAV Based Imaging Platform for Monitoring Maize Growth Throughout Development. bioRxiv Prepr. doi:http://dx.doi.org/10.1101/794057. The.

Torres-Sánchez, J., López-Granados, F., De Castro, A. I., and Peña-Barragán, J. M. (2013). Configuration and Specifications of an Unmanned Aerial Vehicle (UAV) for Early Site Specific Weed Management. PLoS One 8. doi:10.1371/journal.pone.0058210.

Torres, A., and Pietragalla, J. (2012). "Crop morphological traits," in Physiological Breeding II: A Field Guide to Wheat Phenotyping, eds. A. J. D. Pask, J. Pietragalla, D. M. Mullan, and M. P. Reynolds (Mexico D.F: CIMMYT), 1–30. doi:10.1017/CBO9781107415324.004.

Tshikunde, N. M., Mashilo, J., Shimelis, H., and Odindo, A. (2019). Agronomic and Physiological Traits, and Associated Quantitative Trait Loci (QTL) Affecting Yield Response in Wheat (Triticum aestivum L.): A Review. Front. Plant Sci. 10, 1–18. doi:10.3389/fpls.2019.01428.

Ullman, S. (1979). The interpretation of structure from motion. Proc. R. Soc. Lond. B. Biol. Sci. 203, 405–426. doi:10.1098/rspb.1979.0006.

USDA (2018). World Agricultural Outlook Board.

Vautherin, J., Rutishauser, S., Schneider-Zapp, K., Choi, H. F., Chovancova, V., Glass, A., et al. (2016). PHOTOGRAMMETRIC ACCURACY and MODELING of ROLLING SHUTTER CAMERAS. in ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences (Copernicus GmbH), 139–146. doi:10.5194/isprs-annals-III-3-139-2016.

Walter, A., Liebisch, F., and Hund, A. (2015). Plant phenotyping: From bean weighing to image analysis. Plant Methods 11, 1–11. doi:10.1186/s13007-015-0056-8.

Walters, J. P., Archer, D. W., Sassenrath, G. F., Hendrickson, J. R., Hanson, J. D., Halloran, J. M., et al. (2016). Exploring agricultural production systems and their fundamental components with system dynamics modelling. Ecol. Modell. 333, 51–65. doi:10.1016/j.ecolmodel.2016.04.015.

Wang, X., Singh, D., Marla, S., Morris, G., and Poland, J. (2018). Field-based high-throughput phenotyping of plant height in sorghum using different sensing technologies. Plant Methods 14, 1–16. doi:10.1186/s13007-018-0324-5.

Watanabe, K., Guo, W., Arai, K., Takanashi, H., Kajiya-Kanegae, H., Kobayashi, M., et al. (2017). High-throughput phenotyping of sorghum plant height using an unmanned aerial vehicle and its application to genomic prediction modeling. Front. Plant Sci. 8, 1–11. doi:10.3389/fpls.2017.00421.

Westoby, M. J., Brasington, J., Glasser, N. F., Hambrey, M. J., and Reynolds, J. M. (2012). "Structure-from-Motion" photogrammetry: A low-cost, effective tool for geoscience applications. Geomorphology 179, 300–314. doi:10.1016/j.geomorph.2012.08.021.

Wilke, N., Siegmann, B., Klingbeil, L., Burkart, A., Kraska, T., Muller, O., et al. (2019). Quantifying Lodging Percentage and Lodging Severity Using a UAV-Based Canopy Height Model Combined with an Objective Threshold Approach. Remote Sens. 11, 515. doi:10.3390/rs11050515.

Wilks, S. S. (1938). The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. Ann. Math. Stat. 9, 60–62. doi:10.1214/aoms/1177732360.

Willkomm, M., Bolten, A., and Bareth, G. (2016). Non-destructive monitoring of rice by hyperspectral in-field spectrometry and UAV-based remote sensing: Case study of field-grown rice in North Rhine-Westphalia, Germany. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch. 2016-Janua, 1071–1077. doi:10.5194/isprsarchives-XLI-B1-1071-2016.

Wolak, M. (2018). nadiv:(Non)Additive Genetic Relatedness Matrices. 2.16.0.0.

Würschum, T., Langer, S. M., and Longin, C. F. H. (2015). Genetic control of plant height in European winter wheat cultivars. Theor. Appl. Genet. 128, 865–874. doi:10.1007/s00122-015-2476-2.

Yang, B., Wang, M., Sha, Z., Wang, B., Chen, J., Yao, X., et al. (2019). Evaluation of aboveground nitrogen content of winter wheat using digital imagery of unmanned aerial vehicles. Sensors (Switzerland) 19, 1–18. doi:10.3390/s19204416.

Yang, G., Liu, J., Zhao, C., Li, Z. Z., Huang, Y., Yu, H., et al. (2013). Unmanned Aerial Vehicle Remote Sensing for Field-Based Crop Phenotyping: Current Status and Perspectives. Front. Plant Sci. 126, 867–887. doi:10.1007/s00122-013-2066-0.

Yeyin, S., Neely, L., Thomasson, J., Murray, S., Pugh, N., Rooney, W., et al. (2016). Unmanned Aerial Vehicles for High- Throughput Phenotyping and Agronomic. PLoS One c, 1–26. doi:10.5061/dryad.65m87.

Yuan, W., Li, J., Bhatta, M., Shi, Y., Baenziger, P. S., and Ge, Y. (2018). Wheat height estimation using LiDAR in comparison to ultrasonic sensor and UAS. Sensors (Switzerland) 18. doi:10.3390/s18113731.

Yue, J., Yang, G., Li, C., Li, Z., Wang, Y., Feng, H., et al. (2017). Estimation of winter wheat above-ground biomass using unmanned aerial vehicle-based snapshot hyperspectral sensor and crop height improved models. Remote Sens. 9. doi:10.3390/rs9070708.

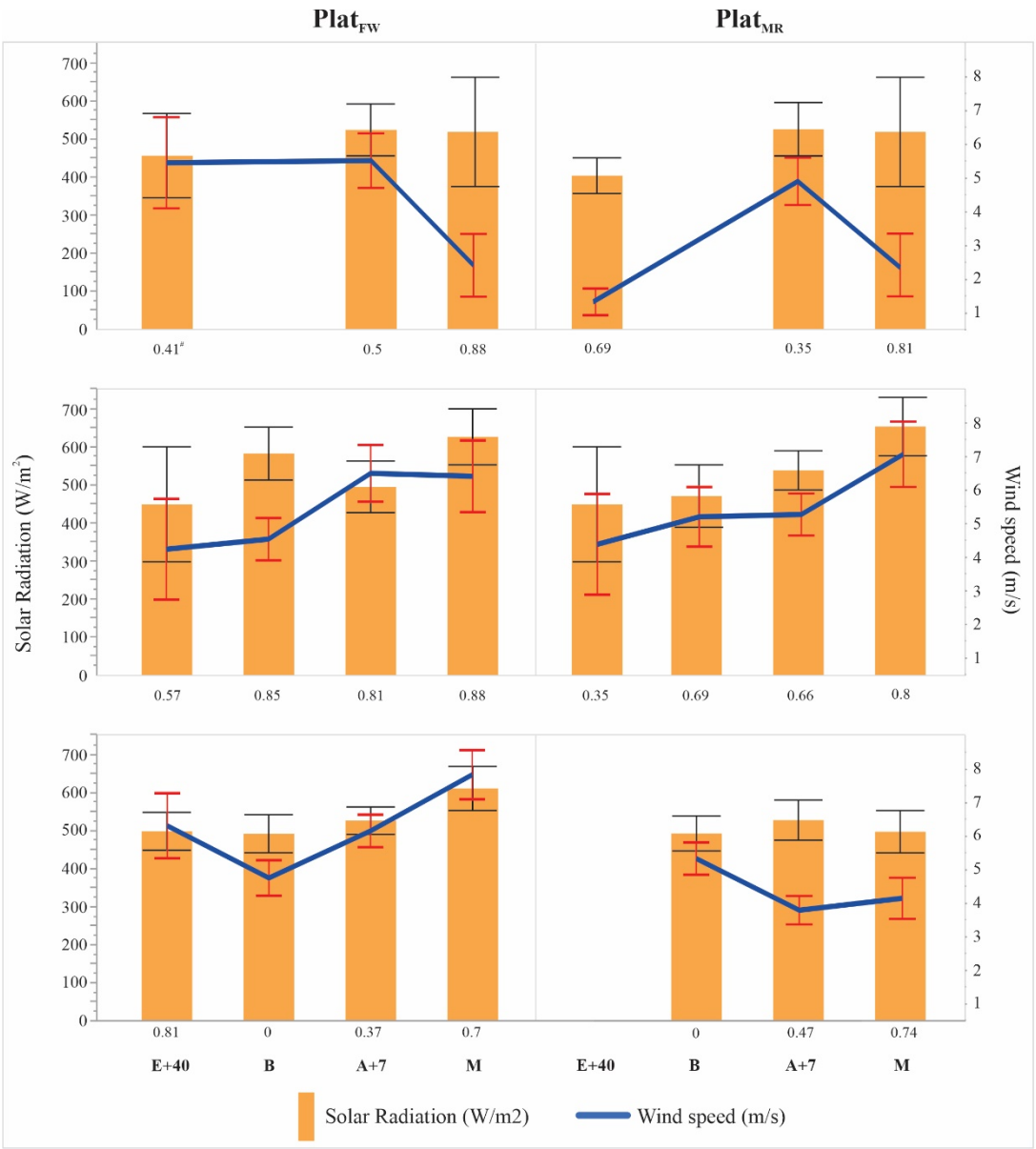
Zadoks, J. C., Chang, T. T., and Konzak, C. F. (1974). A Decimal Code for the Growth Stages of Cereals. Weed Res. 14, 415-421.

Zhang, H., Sun, Y., Chang, L., Qin, Y., Chen, J., Qin, Y., et al. (2018). Estimation of grassland canopy height and aboveground biomass at the quadrat scale using unmanned aerial vehicle. Remote Sens. 10. doi:10.3390/rs10060851.

Zhao, C., Zhang, Y., Du, J., Guo, X., Wen, W., Gu, S., et al. (2019). Crop phenomics: Current status and perspectives. Front. Plant Sci. 10. doi:10.3389/fpls.2019.00714.

Zhou, L., Gu, X., Cheng, S., Yang, G., Shu, M., and Sun, Q. (2020). Analysis of plant height changes of lodged maize using UAV-LiDAR data. Agric. 10. doi:10.3390/agriculture10050146.

8. Supplemental Material



Supplementary Figure 1. Average of flight environment conditions for two breeding cycles around the time of flight campaign at 40 days after emergence (E+40), booting (B), seven days after anthesis (A+7), and maturity (M). The black and red cross lines represent, respectively, the 95% confidence interval computed for solar radiation and wind speed. #R-squared (R^2) value between PHs ground-truth measurements (PHground) and UAS-based estimations (PHaerial).

Supplementary Table S1. Pix4D processing parameters for generating the Point Cloud Densification and final DSMs details for fixed-wind ($Plat_{FW}$) and multi-rotor ($Plat_{MR}$) platforms.

	HiBAP-I Y17		**HiBAP-II Y18**		**HiBAP-I Y18**	
	$Plat_{FW}$	$Plat_{MR}$	$Plat_{FW}$	$Plat_{MR}$	$Plat_{FW}$	$Plat_{MR}$
Keypoints image scale	1/2	1/2	1/2	1/2	1/2	1/2
Image Scale Quality of Point Cloud Densification	1/2	1/4	1/2	1/4	1/2	1/4
Average Number of 3D Densified Points[#]	3385701	7024707	5716299	4155111	1991084	6565985
Average Density of point cloud (per m^3)[$]	569.73	3950.22	1566.28	2697.50	1739.54	2423.69

[#]and [$]are the total number of 3D densified points obtained, and the average number of 3D densified points obtained for the project per cubic meter, respectively, across all projects performed in this study.

CHAPTER 2

OPTIMIZATION OF TEMPORAL UAS-BASED IMAGERY ANALYSIS TO ESTIMATE PLANT MATURITY DATE FOR SOYBEAN BREEDING

VIÇOSA - MINAS GERAIS

2020

ABSTRACT

Measuring the days from sowing to physiological maturity of each plot is a necessary and ubiquitous feature of soybean variety trials. The main drawback of visually dating soybean maturity is the sheer scale of note recording entailed and the frequency at which these notes need to be taken. The overall aim of this study was to build upon prior work in using low-cost UAS-based RGB cameras to estimate soybean maturity date by examining the effect of different vegetation index values, extraction methods from each plot's region of interest, different statistical models, and flight frequencies. Maturity dates collected from five environments with 53 experimental trials (4415 plots) were visually dated, and aerial RGB images were collected. Using the mean greenness leaf index on each plot combined with LOESS regression, we achieved high correlations between ground and UAS-based estimates (r = 0.84 – 0.97). Precision, quantified by the trial broad-sense heritability, was greater for UAS-based dates in 29 of 53 trials, and nearly equivalent in 11 more trials. We found that 54% of the significant deviations between ground and UAS-based estimates were caused by inaccurate UAS-based estimates, while 46% were caused by errors in the ground-based estimates. Reasons for these inaccurate estimates were attributed to lodging, presence of weeds, low germination, and within-line genetic heterogeneity in the plots. A detailed description of the analysis pipeline is provided, along with a user-friendly R script to extract individual plot maturity date estimates from orthomosaic images.

Keywords: Maturity pipeline; RGB camera; High Throughput Phenotyping; Drone imagery; Senescence; *Glycine max*; Vegetation Index.

RESUMO

Medir os dias desde a semeadura até a maturidade fisiológica de cada parcela é uma avaliação necessária e essencial nos testes de variedades de soja. A principal desvantagem de datar visualmente a maturidade de soja se deve a simples escala de mensurações envolvidas e a frequência com que essas mensurações precisam serem tomadas. O objetivo geral deste estudo foi elaborar um esquema de trabalho para estimar a data de maturação de soja por meio do uso de câmeras RGB acopladas ao sistema aéreos não tribuláveis (UAS) de baixo custo. Com base em trabalhos anteriores, o presente estudo visa examinar o efeito de diferentes valores de índices de vegetações, métodos de extração em cada região de interesse da parcela, diferentes modelos estatísticos, e frequências de voo. As datas de maturidade coletadas em cinco ambientes com 53 ensaios experimentais (4415 parcelas ao total) foram datadas visualmente, e imagens aéreas RGB foram coletadas. Usando o índice de folha esverdeado em cada parcela combinado com a regressão LOESS, alcançamos altas correlações entre as estimativas baseadas em UAS e terrestres (r = 0,84 - 0,97). A precisão, quantificada pela herdabilidade no sentido amplo, foi maior para as datas utilizando UAS em 29 de 53 ensaios e quase equivalente em outros 11. Encontramos que em 54% das grandes diferenças entre as estimativas utilizando UAS e terrestres foram causados por estimativas de UAS imprecisas, enquanto 46% foram causados por erros nas estimativas baseadas em avaliações a nível de solo. As razões para essas estimativas imprecisas foram atribuídas ao acamamento, presença de ervas daninhas, baixa germinação e heterogeneidade genética intrínseca nas parcelas. O presente trabalho também visa fornecer uma descrição detalhada do pipeline para estimar data de maturidade, juntamente com um script de fácil uso no software R para extrair as estimativas de cada parcela individual utilizando imagens de orthomosaicos.

Palavras-chave: Pipeline para maturidade de plantas; Câmera RGB; Fenotipagem de Alto Rendimento; Imagens de drones; Senescência; *Glicina max*; Índice de Vegetação.

1. Introduction

Measuring the days from sowing to physiological maturity of each plot is a necessary and ubiquitous feature of soybean variety trials, both in breeding programs and commercial variety testing programs. Soybean varieties vary in their time from germination to physiological maturity, which is controlled by genetic factors that respond to both photoperiod and temperature (Setiyono et al., 2007; Schoving et al., 2020). Collecting reliable data on days to maturity for each variety is essential for determining proper adaptation to a particular latitudinal zone. Varieties that mature too late will be damaged by frost, while those that mature too early may not maximize yield in a latitudinal zone. Another key reason for the collection of maturity dates at the field plot level is to inform the comparison of yield between varieties. Varieties that mature later can be expected to have higher yield than those that mature earlier, and therefore yields should only be compared between varieties that mature on similar dates (Mourtzinis and Conley, 2017; Zdziarski et al., 2018).

Soybean breeders record the date of soybean maturity by visually observing when 95% of the pods in a plot reach their mature color (Fehr and Caviness, 1977). The main drawback of visually dating soybean maturity is the sheer scale of note recording entailed and the frequency at which these notes need to be taken. A typical soybean breeding program could quickly require visual maturity dating on tens of thousands of plots spread across multiple locations, with the frequency of dating being two times per week until all plots have reached maturity. This level of phenotyping imposes a high cost in human resources. Moreover, human errors and subjectivity can unnecessarily reduce the reliability of maturity date data.

Unmanned aircraft systems (UASs) have become less expensive and more versatile in recent years, allowing a large variety of cameras to be flown to collect aerial imagery at high spatial and temporal resolution (Torres-Sánchez et al., 2013; Peña et al., 2015; Holman et al., 2016; Araus et al., 2018; Reynolds et al., 2020). Many different applications of UAS technology to improve soybean breeding programs have been studied for traits including soybean abiotic stress tolerance, canopy coverage, maturity date, and yield (Bai et al., 2016; Yu et al., 2016b; Castelao Tetila et al., 2017; Gao et al., 2018; Huang et al., 2018; Jarquin et al., 2018; Dobbels and Lorenz, 2019). The success of any image collection and analysis pipeline depends on various factors, including the choice of a suitable analysis algorithms and statistical models (van Eeuwijk et al., 2019); temporal and spatial resolution of imagery and temporal imagery (Holman et al., 2016; Belgiu and Stein,

2019); and sensor type and treatment of reflectance data (Yang et al., 2017; Chawade et al., 2019; Maes and Steppe, 2019; Krause et al., 2019). Some studies used hand-held devices (Ma et al., 2001; Christenson et al., 2016; Lindsey et al., 2020), which hold the potential to increase throughput and reduce human errors, but these devices still require a great deal of manual measurements, which can be limiting when thousands of plots need to be evaluated. Yu et al. (2016) used UAS-collected imagery combined with a random forest statistical model and found that they could correctly classify over 95% of the plots within one flight date of the ground-based maturity date estimate. However, the binary maturity classification (mature versus not mature on each flight date) used in this study means that the resolution of dating was limited to the flight frequency, which ranged from one flight every three to six days. Using this methodology, achieving daily resolution would mean that a field would need to be imaged each day. Another approach to predicting maturity is to extract many features from an image taken with a multispectral camera, including changes in reflectance values between flight dates, and use techniques in dimension reduction and regression to estimate the maturity date from the extracted image features. Zhou et al. (2019) used this approach by extracting features from imagery collected with a five-band multispectral sensor and associated these predictor variables with maturity date in a calibration set using partial least-squares regression. This approach allowed maturity date estimates to take on any value in between flights. Using a 10-fold cross-validation, UAS-based maturity estimates explained 81% of the variation in ground-based estimates. A downside of this approach is the need to collect ground data on a calibration set at each field trial site, the need for a multispectral sensor, and the overall added complexity of feature extraction and optimization of complex multivariate models.

An ideal method for estimating the maturity date of soybean plots using UAS imagery would use simple red-green-blue (RGB) cameras and simple modeling relating the spectra to maturity date, not requiring calibration. Narayanan et al. (2019) accomplished this by using reflectance data obtained from an RGB camera attached to a UAS, calculated a “green excess index” (NGE), and modeled the decay of the NGE values of each plot over time using a segmented (or piece-wise) linear regression model. Using this approach, Narayanan et al. (2019) found they could estimate soybean plot maturity with good reliability ($R2 = 0.58 – 0.96$) and precision. The broad-sense heritabilities of the UAS-based estimates were comparable to those of the ground-based estimates across 22 trials. However, Narayanan et al. (2019) did not systematically investigate many variables that could affect the accuracy and precision of estimates, such as flight frequency, type of vegetative index, and method of pixel extraction. Moreover, the influence of

various plot characteristics on the accuracy of UAS-based maturity date estimates has not been fully explored and quantified.

The overall aim of this study was to build upon the work of Narayanan et al. (2019) in using low-cost UAS-based RGB cameras to estimate soybean maturity date by examining the effect of (1) three different vegetation index (VIs) values calculated from RGB bands, (2) five extraction methods from each plot's region of interest (ROI), (3) three different statistical models, and (4) four flight frequencies. In addition to studying the effect of these factors on UAS-based maturity date estimates, we also aimed to categorize and quantify underlying reasons for large deviations between UAS-based (DPMaerial) and ground-based (DPMground) maturity dates. Results from this study will inform researchers and breeders on the best practices for UAS-based maturity date estimation not only in terms of UAS practices, but also in terms of field trial design and plot technique. Another goal of ours was to provide a detailed description of methodology and make available an R script that can be deployed by the users.

2. Material and methods

2.1. Experimental design and ground-based dating of plant maturity

A set of 53 trials of breeding lines at different stages of testing in the University of Minnesota (UMN) soybean breeding program were visually dated for maturity and aerially imaged via UAS. The experiments were carried out throughout two growing seasons (2018 and 2019) at three locations in southern Minnesota, U.S. – Waseca (WA) (44°03'56.0"N, 93°33'01.8"W), Lamberton (LA) (44°16'57.0"N, 95°16'09.6"W), and Rosemount (RO) (44°42'33.6"N, 93°04'28.4"W). Individual trials consisted of breeding lines and check cultivars of similar relative maturity so that the total range of maturity within a trial did not exceed one full maturity group. Across trials, maturity groups ranged from late MG 0 to early MG II cultivars and breeding lines. Field plots (experimental units) consisted of two- or four-rows 3.66 m in length and 0.76 m apart. Alleys between the plots were 0.91 m wide. Plots were arranged in a randomized complete block design (RCBD) with two or three replications, depending on the stage of the trial.

Locations were visited for ground-based visual dating of maturity (DPMground) as soon as the earliest maturing varieties in a given location began to senesce. A plot was determined to be

mature when approximately 95% of pods exhibited their mature color (Fehr and Caviness, 1977). DPMground notes were taken by observing and revisiting plots as they matured, eventually assigning a date of maturity to each plot. Plots were visited every three days, on average, and maturity dates were interpolated when it was clear a plot reached maturity in between site visits. The maturity date was expressed as the date of physiological maturity (DPM), which was calculated as the number of calendar days between August 31st and the date of maturity.

2.2. UAS platform and image data collection

DPMaerial was collected using either a DJI Phantom 3 Pro (DJI Technology Co., Ltd., Shenzhen, China) or DJI Inspire 1 model T600 (SZ DJI Technology Co., Shenzhen) equipped with digital red, green, blue (RGB) cameras for both platforms (Zenmuse X3 equipped to Inspire 1 platform). Flights were conducted across the same time range as the visual DPMground was recorded. In 2018, flights were performed once per week, while in 2019, three times per week, or approximately every two to three days until all plots reached maturity. This allowed the simulation of one (1FlyW) or two (2FlyW) flights per week through the masking of individual flights (Figure 1). The 1FlyW scenario was created in two different ways as the flights did not evenly stagger within one week (Figure 1). Differences in DPMground - DPMaerial correlations were observed between these two ways to achieve 1FlyW and thus correlations were simply averaged for this scenario. The number of days between field visits was sometimes dictated by weather conditions (wind and precipitation) and conducted within one hour of solar noon to limit shadow effects.

All flight missions were conducted at an altitude of 61 m with a UAS speed of 5 m/s and images captured with 75% end and side lap with area covered higher than the actual trials area to ensure accurate orthomosaics. As a result, the images had a ground sampling distance (GSD) of 2.4 ~ 2.6 cm. Waypoints and flying routes were automatically generated using the flight planning software, Pix4Dcapture (v4.8.0; Pix4D, Prilly, Switzerland). A tablet was used to prepare missions prior to going to the field, and then uploaded to the main control board on the UAS before each flight. During each flight, the UAS traveled a predetermined route via the software's autonomous flying mode.

Painted red blocks (30 x 30 cm) were placed as permanent ground control points (GCPs) to ensure the georeferencing accuracy of images collected by the UAS platforms. The GCPs were

placed randomly throughout the field location, covering the entire area of interest, and remained in the field for the duration of the season. The GCPs were surveyed using a Garmin eTrex 30x hand-held device (Garmin, Kansas, USA). The experimental and technical details, as well as summarizes the major mission flight specifications for each environment in this study, are shown in Table 1.

Table 1. Details on each environment, unmanned aircraft systems (UASs), and flight settings used in this study. A total of five environments were used between 2018 and 2019 to estimate days of plant maturity (DPMaerial). Each field site varied in area (ha), number of trials, the total number of plots, planting date, number of ground control points (GCPs), and type of unmanned aircraft system (UAS) platforms used.

	2018		**2019**		
	WA	**LA**	**WA**	**LA**	**RO**
Number of trials	12	11	6	9	15
Number of plots	873	794	482	684	1382
Planting date	May 17th	May 16	May 15	May 16	May 26
Area total (ha)	3.04	3.40	2.42	2.14	3.61
Number of flights	9	12	18	18	16
Flight period start/end date	Aug 08 / Oct 12	Aug 15 / Oct 11	Aug 30 / Oct 14	Aug 28 / Oct 16	Aug 28 / Oct 07
GCP number	5	4	4	4	9
UAS Platform	DJI Phantom 3 Pro	DJI Phantom 3 Pro	DJI Phantom 3 Pro	DJI Phantom 3 Pro	DJI Inspire 1

Environmental and management details of this area also are given in Soybean Field Crop Trials Results (link: https://www.maes.umn.edu/publications/field-crop-trials/soybean)

2.3. Image data processing

A pipeline software tool and scripts for data extraction and analysis were developed to accomplish the activities ranging from image capture to statistical analysis of extracted features. The DPMaerial scripts and processes used to perform the image analyses and trait extract are available at https://github.com/volpatoo/HTP-via-drone-imagery. Figure 1 highlights the major steps in this pipeline.

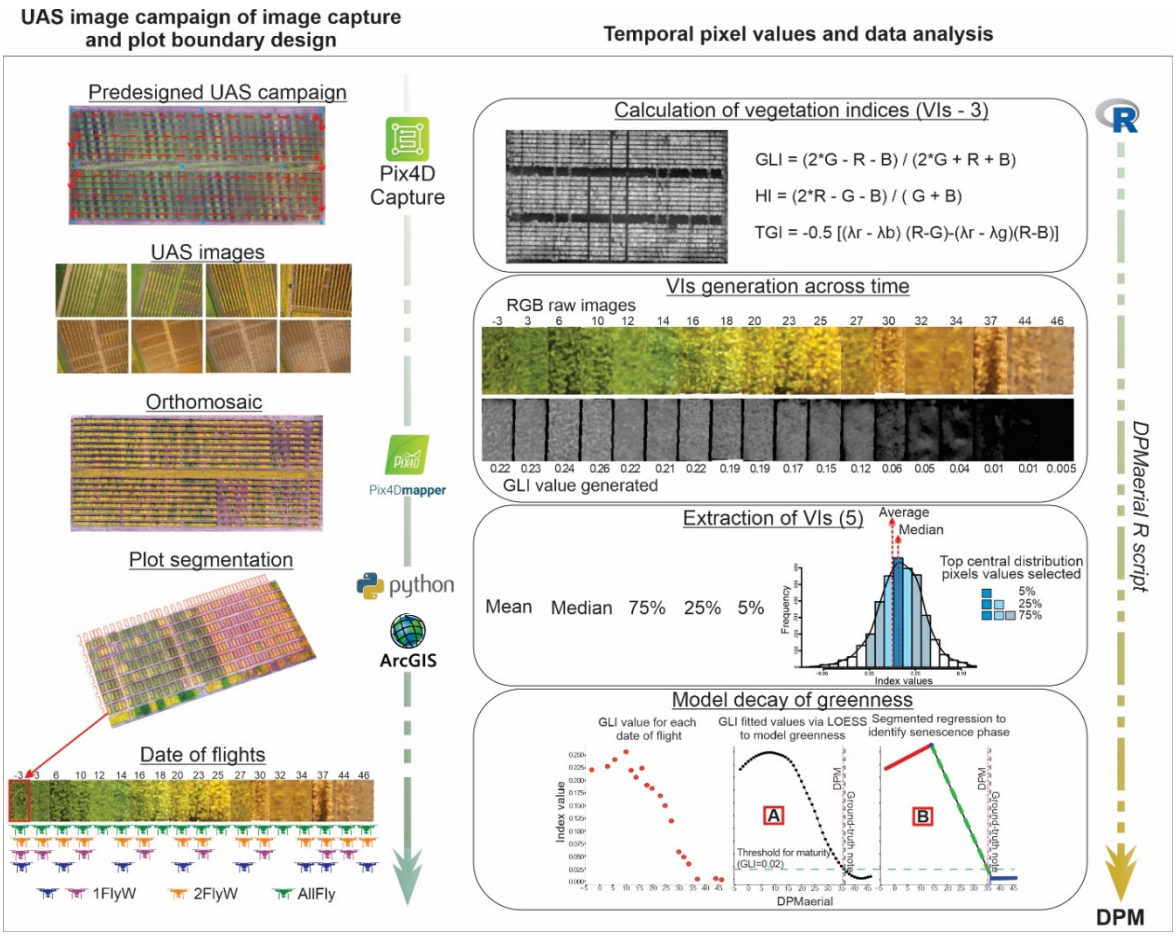


Figure 1. Image capture and analysis pipeline for days of plant maturity (DPM) estimation. The left side shows the UAS flight mission and image process accompanied by the corresponding software used at each step. The right side shows all the criteria to set up the dataset to plant maturity outputs into the R software package according to the levels of factors used: three different vegetation indices (VIs), five extractions pixels extraction methods, and the two statistical methods to optimize the DPM. Local regression (A) and segmented regression (B) models. GLI, greenness leaf index; HI, hue index; TGI, triangular greenness index; VIs, vegetation indices.

Orthomosaics were generated using Pix4D Mapper® software (v4.5.4; Pix4D, Prilly, Switzerland), after the image collection, which provides an automated pipeline including the steps of photo alignment, matching, and bundle adjustment. The WGS 84 datum was used with a projected coordinate system of UTM zone 15 N. Images were imported into Pix4D and then optimized and matched using the default processing options template ‘Ag RGB’ for generating georeferenced orthomosaics. This option generates orthomosaics from overlapping nadir images and outputs a full resolution GeoTIFF file. The GCPs were input and identified into the Pix4D project using the basic manual editor before initial processing.

Python 2.7.1 (Python Software Foundation) software was used to generate the polygon shapefiles (ROI) according to plot boundary delimitation following Haghighattalab et al. (2016) workflow by using the soybean trials imagery before the full flowering stage (R2). ArcGIS (version 10.7, Esri Inc., Redlands, USA) buffer tool was used for polygon adjustments of the ROI to extract vegetation indices procedures. The spatial polygon vectors (or shapefiles) were outlined with a buffer zone of 10 cm for each plot to avoid the influence of neighborhood plots. Additionally, the ROI of the plots were manually aligned, as needed on the center of the two or four rows of every plot in all locations.

Seventeen VIs were calculated using the RGB-bands from images collected in 2018 using the package 'raster' (Hijmans, 2020) in R (Development Core Team, 2018). VIs Geo-TIFF files generated were imported into R for polygon value extractions using the mean pixel values from each ROI. The plot boundary delimitation was performed using the package 'rgdal' (Bivand et al., 2020) to import the shapefile, and the package 'raster' was implemented to manipulate spatial data (pixels) in raster format file, both in R software.

A preliminary analysis was performed using the randomForest R package (Liaw and Wiener, 2018) to relate the 17 VIs to DPMground. The DPMground data was treated as a character, and 70% of the data, at random, was used to train the models, and 30% of the data, at random, was used for validation. The evaluation of the importance of each variable (17 VIs) in the prediction process suggested three indices based on variable importance rankings (Strobl et al., 2007) (data not shown): GLI, TGI (triangular greenness index) and HI (hue index). The GLI index was initially created for use with a digital RGB camera to measure crop cover (Louhaichi et al., 2001) and the TGI to estimate the area of a triangle bounding a leaf reflectance spectrum (Hunt et al., 2012). Both indices also have been applied in studies to measurements of differences in leaf chlorophyll content at leaf and canopy scales. The HI index has also been used in vegetation applications (Escadafal et al., 1994; Zhou et al., 2019). These three indices were subsequently used to model the decay in plot greenness through the senescence phase in this study.

The value of each index corresponding to plant maturity (i.e., threshold) was calculated by taking the average pixel values from each VIs of all plots that were visually scored as mature on the day of imaging (see Supplementary Figure S1 for distribution of index values for each of the three indices at plot maturity). The effect of the predetermined threshold at which estimated maturity was evaluated using the correlation between DPMground and DPMaerial across different

VI thresholds. Thus, testing different ranges for each VI threshold, as described by Narayanan et al. (2019), the best ones based on the highest correlation were selected as the optimized threshold (results not shown).

Using only the top three VIs selected (GLI, HI, and TGI), the same procedures were implemented to manipulate pixel value extractions from the entire dataset with the packages 'rgdal' and 'raster' into the R software. Five different pixel extraction procedures were performed by extracting the central 75th, 25th, and 5th percentile (75, 25, and 5% selected of top central distribution pixels values), the mean pixel values, and median pixel values inside the ROI from each plot. The purpose of testing different pixel extraction selections is to test if they can be used in an attempt to avoid off-target pixels (e.g., weeds, plants of neighboring plots, or even patches of bare ground between plants).

2.4. Models to estimate days to plant maturity

Three models to estimate DPMaerial from RGB spectra were investigated: non-parametric local polynomial regression (LOESS), segmented linear regression (SEG), and random forest classification (RF). LOESS is a localized regression approach that fits simple linear models to localized subsets of data determined by a nearest neighbor algorithm. In this study, the default settings of the LOESS function were implemented using the '*stats'* package from the base R software in order to interpolate VIs values between flights following the methodologies described in Cleveland and Devlin (1988).

DPMaerial based on decaying canopy greenness using SEG was assessed similar as was done by Narayanan et al. (2019). SEG is a form of regression that allows multiple linear models to be fit to the data for different ranges of the explanatory variable (x), which was days after August 31st in this study. We used the '*segmented'* package (Muggeo, 2019) in R to identify the senescence phase curve. The *segmented* function first tests for differences between slopes using the Davies test (Davies, 1987), and then estimates the break-points between the linear models using maximum likelihood (Muggeo, 2003). Break-points are the values of x where the linear slope of the function changes. When the number of flights was equal to or greater than two per week, two break-points were modeled, resulting in three curves. Otherwise, only one break-point was modeled because of difficulty in converging on solutions when two break-points were attempted to model one flight

per week. The curve with the greatest absolute value in slope was selected to automate the selection of the curve representing the senescence phase of each plot. The scales of the indices were such that senescence phase curve slopes were negative for GLI and TGI and positive HI. Using the senescence phase curve, the estimated DPMaerial was calculate as $\frac{(T_i - a_{(ij)})}{b_{(ij)}}$ where T_i is the threshold for the VI *i* (refer to Supplementary Figure 1); $a_{(ij)}$ is the intercept of the senescence phase curve model for VI *i* and plot *j*, and $b_{(ij)}$is the slope of the senescence phase model of VI *i* and plot *j*. The function *'seg.control'* from the *'segmented'* package was used as an auxiliary function to fitting the model using 50 bootstrap samples.

The DPMaerial script works through of inputs parameter controlled by a control file (control.txt), which is processed directly by the "matuRity_est.R" file, which also loads the libraries and custom functions in "functions_PM.R.". As an input requirement, the georeferenced orthomosaics were generated on the Pix4D across multiple time points, and the shapefiles to ROI delimitation were created in Python and tuned using ArcGIS. The R script first converts each orthomosaic into a user-defined RGB based index (GLI, HI, TGI) and extracts a user-defined quantile (75%, 25%, 5%, mean, or median) of pixel values within the ROI defined by the shapefile for each time point of image capture. Then, the user also can explore the outputs (DPMaerial) either for LOESS or SEG regression models according to a user-defined threshold.

For the assessment of binary classification of maturity using the machine learning approach, mean pixel values of GLI, TGI, and HI were used from each plot in all five environments over 2018 and 2019. The RF algorithm was used for maturity classification with JMP Pro (JMP®, Pro 14, SAS Institute Inc., Cary, NC, 1989–2007) software. One environment was used to test the RF model, while the other four environments were used to train and validate the RF model. The RF used 500 trees in the forest with ten minimum splits per tree and ten multiple fits over the number of terms. The VIs values were associated with the binary recorded maturity variable (mature/not mature based on ground-truth notes) across data points according to the following equation:

$$Maturity\ (yes\ /\ no\ ) \sim GLI\ + TGI\ + HI$$

DPMground was matched with the nearest date for the temporal image data (Table 2 – 'Number of flights') in each location, and the individual plot was classified into mature or not-mature for a given date according to its probability chance. The DPMaeral for the testing set using the RF approach was then compared with the DPMground to measure the prediction performance.

2.5. Data analysis and model validation

Each of the 53 trials included in this study were arranged in a RCBD with at least two replications in each environment. Within each environment, best linear unbiased predictions (BLUPs) of the genotype were calculated for the DPMground and DPMaerial dates. BLUPs were obtained for each trial using the following model:

$$y_{ij} = \mu + g_i + r_j + \varepsilon_{ij}$$

where y_{ij} is the value for the plot of genotype *i* and replicate *j*, μ is the grand mean, g_i is the random effect for genotype *i*, r_j is the fixed effect of replicate *j*, and ε_{ij} is the random residual. Both the genotype and residual effects were assumed to be independent and identically distributed random variables from normal distributions. Genotypic (σ_g^2) and residual (σ_e^2) variance estimates from each trial were used to calculate the broad-sense heritability (H) both for DPMground and DPMaerial given as $H = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}$ (Piepho and Möhring, 2007). Statistical analyses were carried out using ASReml-R (version 4.1) software (ASReml, VSN International, Wallingford, UK).

DPMground and DPMaerial were compared and validated using Pearson correlation coefficients (*r*), coefficients of determination (R^2), root-mean-square errors (*RMSE*), and *H* calculated on individual plot values in each location. Confidence intervals (CIs - 95%) of Pearson correlation coefficients were calculated using the pairwise correlation function in JMP Pro software analysis (JMP®, Pro 14, SAS Institute Inc., Cary, NC, 1989–2007), which implements Fisher's z' transformation. Also, 95% CIs of the *H* were computed using the standard errors (Schweiger et al., 2016) from each trial with the '*pin*' function through the R package '*nadiv'* (Wolak, 2019).

Outliers were determined using two methods *i*) based on DPMground alone and *ii*) by analyzing deviations between DPMaerial and DPMground. For (*i*), outliers were determined using standardized residuals of DPMground by JMP Pro software analysis. An analysis of variance was conducted for each of the 53 trials, including the genotype and replicate sources of variation. Standardized residuals were calculated for the DPMground value of each plot, and individual values were considered as outliers if the absolute value of the standardized residual was greater than three. For (*ii*), a regression model between DPMaerial and DPMground was analyzed using the function '*outlierTest'* from the '*car'* package (Fox et al., 2019) in R software with the default

settings from the package. The outlier test ($P < 0.01$) uses a Bonferroni *p*-value to test for extreme observations between the DPMaerial and DPMground in the linear model.

3. Results

We deployed a low-cost multi-rotor unmanned aircraft system (UAS) equipped with RGB cameras to image 53 soybean variety trials in five environments for a total of 4215 plots. Optimization of image collection and analysis was explored by varying the VIs used to model the decay of plot greenness, the proportion of pixels extracted from each plot image, statistical model, and flight frequency. Accuracy and precision of DPMaerial estimates were evaluated using correlations, regressions, and estimates of heritability. Finally, the nature of the UAS-based outliers was examined in order to improve the practical applicability to DPMaerial estimation.

3.1. Ground-notes for plant maturity

The distribution of DPMground was first analyzed across the range of 53 trials in the five environments spanning 2018 and 2019. Supplementary Figure S2 displays the extensive range in DPMground with soybean plots ranging in maturity from 5 to 50. The range in DPMground varied among trials. The standard deviation within trials ranged from 2.05 to 8.14 in 2018 LA, 1.87 to 6.65 in 2018 WA, 3.56 to 6.27 in 2019 LA, 2.96 to 4.2 in 2019 WA, and 1.79 to 5.4 in 2019 RO. The boxplot Supplementary Figure S2 also shows the potential outliers in each trial across the environments, which at 2019 RO location throughout trials shows more DPMground data points under this condition.

3.2.Vegetation indices and pixel extraction method to estimate plant maturity

The statistical model for DPMaerial estimation across VIs and pixel extraction proportions was studied. Three VIs -- GLI, HI, and TGI -- were calculated from the RGB spectra data to determine which best-modeled decay in canopy greenness for purposes of DPMaerial estimation. Figure 2 displays correlation coefficients between DPMground and DPMaerial for each of the

different levels of the factors VI, pixel extraction proportion (25q, 5q, 75q, mean, and median), and statistical models (LOESS and SEG).

Overall, shaded box in Figure 2, GLI resulted in the best DPMaerial estimations, although differences never exceeded the CI when comparisons were made between VIs within the LOESS model (black lines). However, using the SEG model (green lines), statistical difference according to the CI was observed at HI in 2019 LA and RO. LOESS showed a significative estimation of accuracy compared with SEG in 2018 locations for all VIs deployed except for GLI in LA 2018. These results, combined with those of Narayanan et al. (2019), persuaded us to choose GLI as the single index value to extract from each ROI in all subsequent analyses for LOESS, SEG, and RF models.

The DPMaerial – DPMground correlations for different pixel extraction proportions were also very similar across all environments (Figure 2 in no painted box). Because each of the five extraction values provided similar overall results across uniform full vegetation soybeans plot, the mean value extracted in each ROI was considered in all subsequent analyses.

Overall, the LOESS model showed a slight advantage in 2019 LA and RO and a significant difference against the SEG model 2018 LA and WA. The only exception was in 2019 WA, in which the SEG model showed a higher correlation than the LOESS model, however, the difference did not exceed one CI (Figure 2).

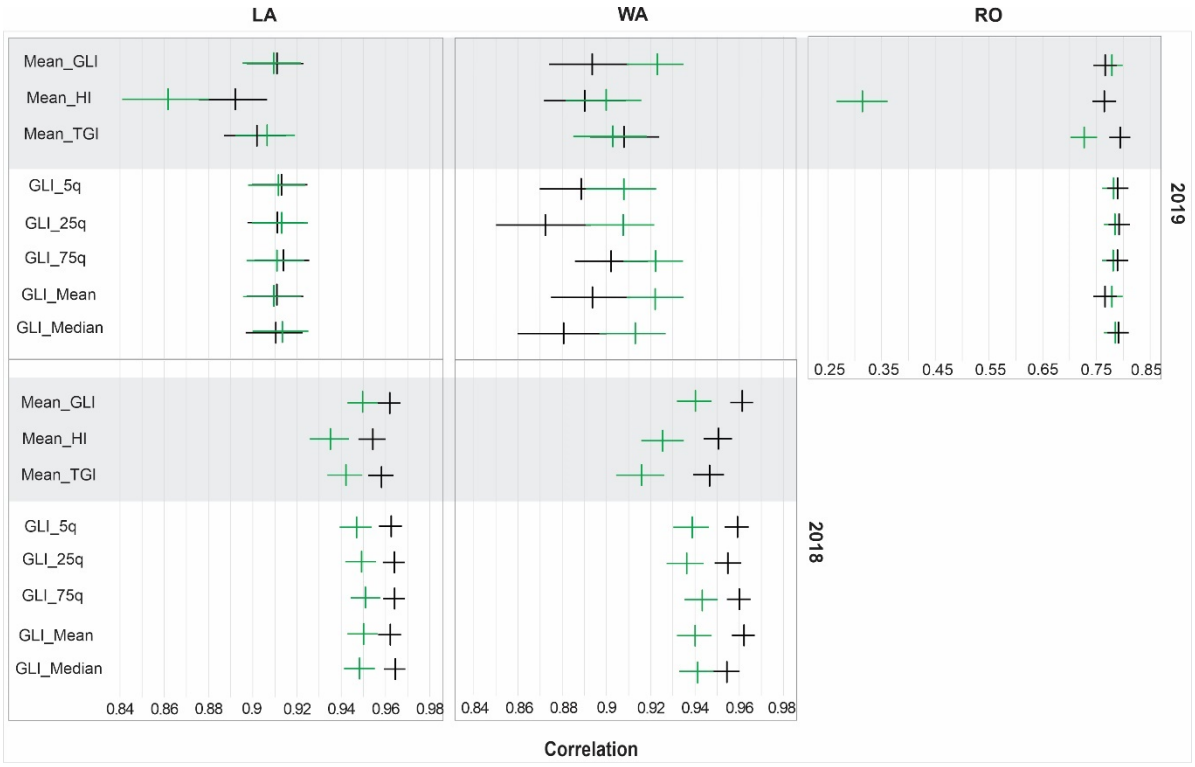


Figure 2. Correlations between ground-truth notes (DPMground) and UAS-based imagery (DPMaerial) for local regression (LOESS, vertical black lines) and segmented regression (SEG, vertical green lines) models and their correspondent confidence intervals (horizontal lines) using all flights available for each environment in 2018 and 2019 crop year. The gray box shows the results for the three different VIs (GLI, TGI, and HI) using only mean value pixels extracted, while the non-colored ones represent the results of the correlation for the comparison of five pixels values extraction methods (mean, median, 75q, 25q and 5q) using only GLI index. LA: Lamberton, WA: Waseca, and RO: Rosemount.

3.3.Frequency of flights and statistical modeling to estimate plant maturity

The flight number per week variable was combined with two statistical models – LOESS and SEG – in determining whether a statistical model was particularly best suited to a specific flight frequency using mean pixel value extraction and GLI vegetation index. Additionally, the performance of the RF model was compared with SEG and LOESS models using all flight campaigns available and the same factors – GLI and mean – (Figure 3).

In examining the ideal number of flights per week for both models (LOESS – solid black lines and SEG – solid green lines in Figure 3), the correlations between DPMground and DPMaerial were similar or non-significative according to CI throughout the settings of simulated flight

frequencies in WA and RO locations. However, the SEG model using one flight per week reached significant correlations values among flight frequency in LA. The RF model performed similar results across locations, except in LA where this approach showed the lowest values of correlations (r = 0.84 RF vs. 0.91 for LOESS and SEG) or the highest deviation compared with LOESS and SEG models (Figure 3) using all flights available (AllFly). As a result, LOESS and SEG models achieved average similar correlations across locations (r = 0.86 LOESS vs. r = 0.85 SEG), and both outperformed the RF model classification approach (r = 0.82) when all flights were used. Furthermore, using the previously variables to compare between locations, WA and LA showed better correlations than RO (r = 0.89 LA and 0.90 WA vs. r = 0.77 in RO).

Therefore, the analogous correlations among the frequency of flights allowed us to choose just one flight per week as the best option for the plant maturity estimation in order to increase the pipeline-throughput feasibility. Finally, this setting allowed inclusion of 2018 data to validation results as flights were only conducted once a week in that year.

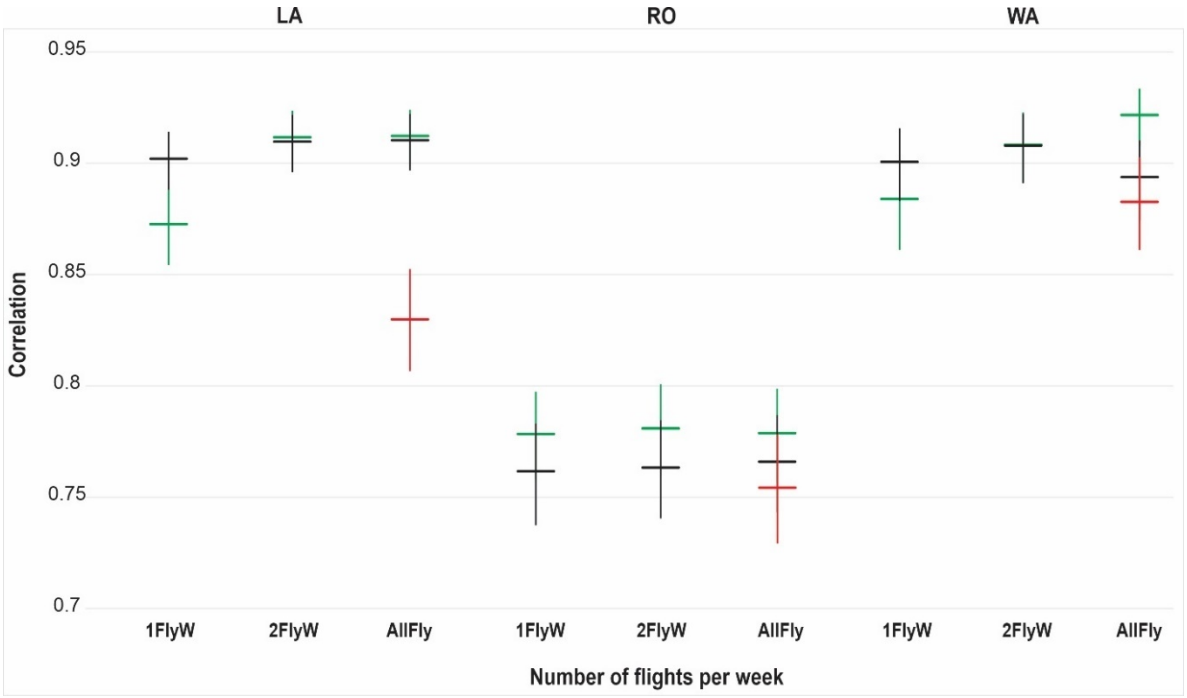


Figure 3. Correlations between ground-truth notes (DPMground) and UAS-based imagery (DPMaerial) for local regression (LOESS, horizontal black lines), segmented regression (SEG, horizontal green lines), and random forest (RF, horizontal red lines) models using GLI index and considering mean pixels values extracted during the temporal flights in 2019. The vertical correspondent lines are the confidence intervals at the 0.95 probability level. 1FlyW: average value from the one flight per week settings; 2FlyW: two flights per week; All: considering all flights available. LA: Lamberton, WA: Waseca, and RO: Rosemount locations.

3.4.Validation strategy

The general accuracy of DPMaerial estimates was assessed by choosing the pipeline settings determined to maximize the correlations per the previously described results above. The chosen settings were one flight per week, GLI vegetation index, LOESS estimation method, and the mean of ROI for pixel extraction. As discussed above, the results from different settings were very similar, and so the chosen settings were also chosen based on their ease and simplicity. Through these settings, the overall relationship between DPMaerial and DPMground was reliable and accurate where the mean of R^2 values was 0.8, ranging from 0.57 to 0.93, *RMSE* ranging from 1.79 to 2.99 across the five environments assessed (Supplementary Figure S3). The lowest values of the R^2 and the highest *RMSE* was found in 2019 RO (R^2 = 0.57 and *RMSE* = 2.99).

The *H* for DPMaerial was numerically higher than the *H* for DPMground for 29 of the 53 experimental trials (55%), and nearly equivalent for 11 more trials (Figure 4). The DPMaerial *H* of many of the RO 2019 trials was markedly better than DPMground, which is a likely reason why the R^2 between DPMaerial and DPMground was substantially lower than those of other environments.

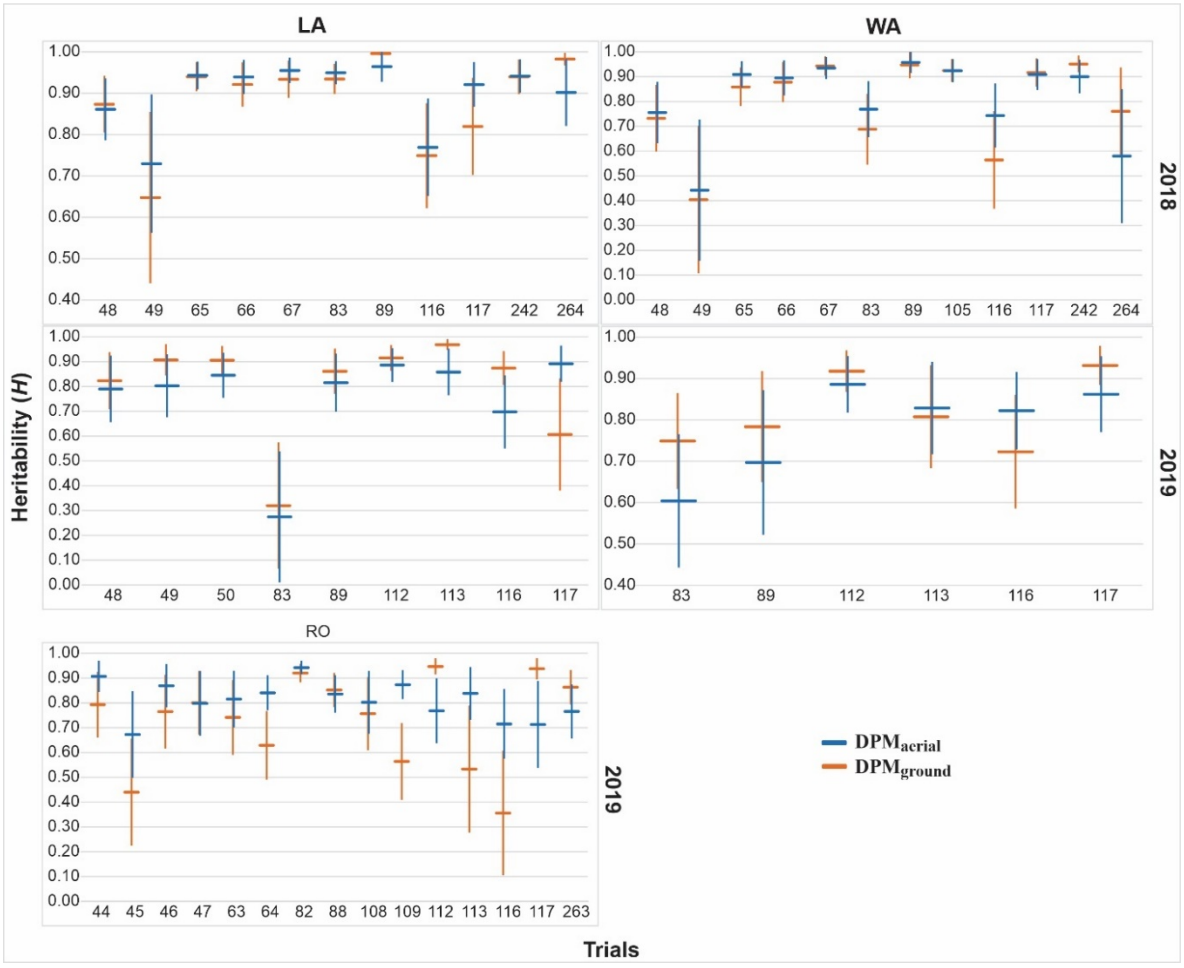


Figure 4. Broad sense heritability (*H*) for 53 trials in five locations comparing UAS-based (DPMaerial, horizontal blue lines) to estimate days of plant maturity via Local regression model from GLI index, average pixels values extracted, and one flight per week (1FlyW) to human collected maturity scores (DPMground, horizontal orange lines). The vertical correspondent lines are confidence intervals at the 0.95 probability level. LA: Lamberton, RO: Rosemount, and WA: Waseca

Finally, the practical issues to potential flag the outliers were first obtained according to the studentized residuals (*i*) for later detection of its statistical significance using the Bonferroni test (*ii*). A total of 212 outliers were identified out of 4,415 observations (~5%) assessed in this study combine both methods (outliers detected with *i* and *ii* criteria). These outliers were further inspected by visual observations of the UAS-collected imagery of the plots to determine the likely cause for the discrepancy between DPMaerial and DPMground. Two broad categories of outliers were observed. For 46% of the outliers, the DPMground was determined to be erroneous based on the re-inspection of the images. That is, the DPMaerial measure looked more appropriate than the DPMground due to be visually closer to the maturity date by the image confirmation. For the other

54% of outliers, the DPMaerial appeared to be erroneous. By further inspecting the images, we determined the reasons for the DPMaerial errors could be attributed mostly (48%) to within-plot heterogeneity, which could be due to seed lot contamination or residual heterogeneity at maturity loci within F4-derived families. Other reasons underlying deviations of DPMaerial included lodging, weeds, and low germination (Figure 5).

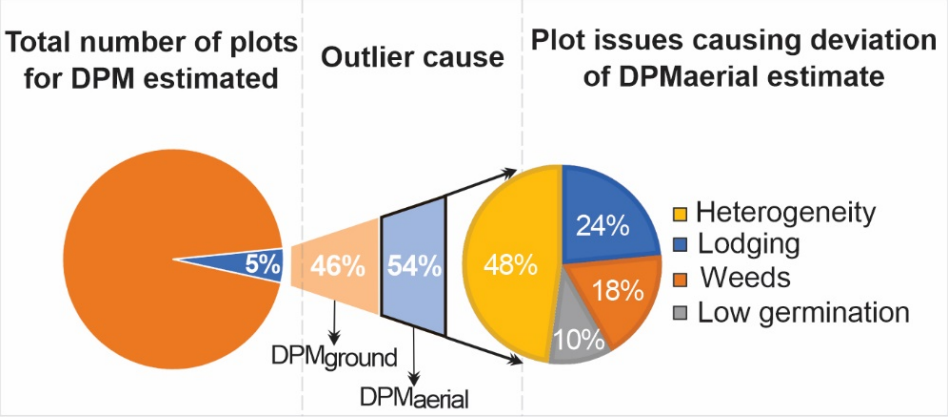


Figure. 5. 212 or 5% outliers flagged according to its specific issues across all locations in 2018 and 2019. The flagged outliers were defined as those that represent a significant deviation between DPMaerial and DPMground.

4. Discussion

4.1.Temporal greenness modeling and flight frequency

Modeling the decay of soybean canopy greenness through remote sensing and statistical models is useful for estimating soybean maturity date. It has also been shown that taking a classification approach using random forest models on individual time points can be accurate (Yu et al., 2016). The challenge with the classification approach, however, is that soybean plots will mature in between flights, and by using a binary classification scheme, it is not possible to interpolate between flights, thus limiting the resolution to the flight frequency. Narayanan et al. (2019) proposed a method to overcome this limitation by modeling greenness decay over time by first characterizing three different phases of each plot and then using regression on the senescence phase to estimate the date of plot maturity. Our study builds upon Narayanan et al. (2019) by exploring different levels of settings in the image collection and maturity analysis in terms of balancing simplicity with accuracy and precision. Another output of this study is the release of a user-friendly script to estimate DPMaerial, allowing the researchers and breeding programs to easily implement these methods.

A threshold for index values is necessary to determine when a plot has reached a level of greenness corresponding to maturity (Figure 1 – A and B). Threshold values were derived by extracting VIs from images of all fully mature plots in which images were collected on the same day of manual scoring (see Supplementary Figure S1). A higher threshold indicates that maturity occurred early while there was still green residue present in the plots. Narayanan et al. (2019) demonstrated that the effect of threshold adjustment for NGE (equivalent to GLI) on accuracy and concluded that threshold values between 0.01 and 0.05 did not affect estimation accuracy. Our results were similar in that the data collected herein suggested that a GLI threshold of 0.02 was optimal. Nevertheless, testing different threshold levels is recommended in each case because it may change based on region and sensor hardware.

We found that a single flight per week (1FlyW) achieved similar accuracy compared to flying as many as two or three times per week. Minimizing flight frequency would ease data collection logistics in terms of travel to sites for image collection, navigating unfavorable weather conditions, and time in processing imagery. When flight frequency was reduced to only biweekly, a decrease in estimation accuracy was observed, although the estimation accuracy was still high at 0.77 as compared to 0.87 for one flight per week (see Supplementary Figure S4). These results suggest it is not necessary to have all three phases of the canopy greenness decay captured to accurately estimate the maturity date, as was reported to be a requirement by

Narayanan et al. (2019). Depending on flight timing, either the pre-senescence or post-senescence phases might not be captured due to logistical concerns within a breeding program. However, as long as the senescence phase is captured, an accurate calculation of the maturity phenotype for both models (LOESS and SEG) can be made.

UAS-estimated maturity dates rely on time series data to fit a smooth or segmented curve. Since LOESS fits linear models to localized subsets of data, flights are required both before and after the date of maturity. Practically, this means that flights will need to continue until every plot reaches maturity. The SEG model, on the other hand, can estimate DPMaerial even if plots do not reach maturity within the range of flight dates. However, accuracy tends to be reduced due to the adjustment lack of the senescence regression model using a few data points to the prediction extrapolation values. The DPMaerial for eight late breeding lines was not estimated using the LOESS model, and previously variables selected to performing the validation analysis. In contrast, the SEG model estimated it with a difference, in average of six days (r = 0.63) when compared to DPMground.

4.2.Index values and pixel extraction methods

Our results corroborate with Yu et al. (2016) study, in which the blue band was present in the index calculations for the most critical VIs selected (GLI, TGI, and HI). One purpose of VIs is to enhance the vegetation signal while minimizing the solar irradiance and soil background effects (Huete et al., 2002; Wang et al., 2017). Some authors have shown the importance of excluding soil pixels when performing analysis of VIs (Thorp and Tian, 2004; Torres-Sánchez et al., 2013; Schirrmann et al., 2016). Masking of the pixels to be selected can also be a recommended tool to control pixel quality (Dobbels and Lorenz, 2019). However, as well as founded by Narayanan et al. (2019) to estimate maturity date in soybean variety trials, our results also not demanded an extra step to pixel masking according to the implemented pipeline for tracking soybean plot senescence through the time.

Similar results from uncalibrated VIs obtained at different sets of flights were found in our study, even when only one flight per week was used (Figure 3). We mimicked one flight per week by choosing different flights each week to create different scenarios, and these correlated very highly with one another in terms of their estimated maturity dates. Thus, the number of flights at various sky and lighting conditions does not seem to affect relative differences in RGB band reflectance values to the point of influencing maturity date estimation

using regression on canopy greenness decay. Moreover, high correlations between DPMground and DPMaerial, combined with similar or greater *H* values for DPMaerial, suggests that avoiding time-consuming spectral calibration will not impact estimation of DPM using UAS imagery, as demonstrated by Narayanan et al. (2019), as well.

The different measures (e.g., mean, median, 5th percentile) applied to filtering pixels that were tested demonstrated different alternatives that could be deployed according to a specific field condition across temporal image capture. This approach has been used by other authors to select target pixels inside each ROI to estimate plant height (Madec et al., 2017; Hassan et al., 2019) and lodging (Singh et al., 2019). Testing different measures applied to pixel extraction is recommended to assist researchers in choosing which is most appropriate for a specific crop, trait, and field design. Haghighattalab et al. (2016) developed an automatic image classification method but pointed out that post-processing and manual adjustments were still necessary on images of wheat plots late in the season. In our study on soybean maturity date evaluated on elite germplasm under uniform field conditions, we found that the mean of the pixel values was just as good as any other measure. However, these results were obtained after precisely defining the ROI using GIS tools (i.e., ArcGIS software) to adjust the polygons to cover each plot manually. Moreover, we gathered early-season imagery (~R1 growth stage) to facilitate the creation of accurate and centralized plot boundaries in order to easily visualize the plot limitations due to less dense canopy coverage in early growth stages. For fine adjustment of UAS-based imagery applications and to reduce non-target pixel influence (i.e., neighborhood plots) in images, methods for pixels values extractions can be explored plus an additional buffer zone inside each target plot, as performed in this study.

4.3. Field conditions and pipeline performance

There are practical issues to consider in field trials that can introduce noise when implementing UAS-based imagery. In this study, significant deviations between DPMaerial and DPMground measures were visually inspected in the images to determine the sources of these differences. As mentioned, human visual scoring methods can involve human bias and mistakes, which consequently decrease correlations between DPMground and DPMaerial (Chapman et al., 2014; Jin et al., 2017; Song and Wang, 2019). Accurate phenotyping is extremely important in experimental validation (Furbank and Tester, 2011; Großkinsky et al., 2015), and it has been a bottleneck when studying HTP technology, including using UAS-based

imagery (Han et al., 2018; Loladze et al., 2019). An enumeration and detailed characterization of these errors would provide insights relevant to the routine use of these methods.

Fifty-four percent of the significant DPMground – DPMaerial deviations were due to errant DPMaerial estimates (Figure 5). Inspection of the images revealed that these deviations could be categorized into four issues affecting the plots: lodging, weeds, low germination, and genetic heterogeneity (Figure 6). In the case of lodging, an adjacent plot is leaning into the ROI, and the greenness value of the ROI remains above the predetermined threshold, effectively delaying the DPMaerial relative to DPMground estimates. Similar to adjacent plots lodging, weeds within a plot also keep the plot greenness above the threshold past the true maturation date. A similar situation occurs with weeds where the presence of green weeds in a plot delays the DPMaerial estimate relative to DPMground. Low germination results in plots not developing a full canopy before senescence. In this case, the plots tended to be classified by the UAS pipeline as mature before they actually were. Finally, heterogeneity can be present because of either seed contamination or residual genetic heterogeneity within a breeding line derived from a single plant that was not fully inbred. In this case, part of the plot would mature differently than the other part, thus introducing errors for DPMaerial estimates. As we investigated in this study, one potential solution to dealing with these issues is to use the top centralized percentile pixels values (75q, 25q, and 5q) to filter out pixels with VIs values that are abnormal compared to the bulk of pixels representing the plot. This method could also be recommended for imaging in soybean trials with a few rows per plot as in single plant rows. However, as demonstrated in this study, an adequate ROI combined with mean pixel value extraction provided reliable results with no advantage seen in other pixel filtering criteria.

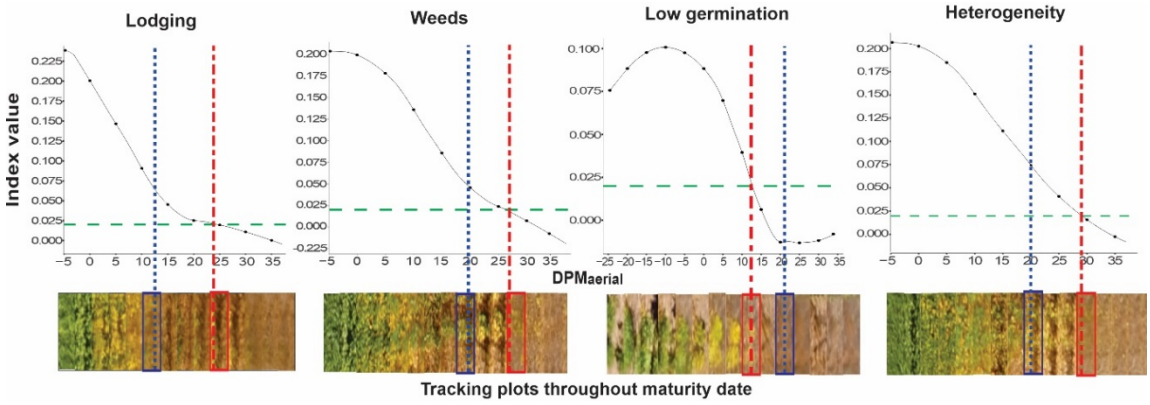


Figure 6. Days to maturity by GLI index values according to common plant breeding issues in random plots across locations using the LOESS model. The blue and red dotted lines represent the ground-truth measurements (DPMground) and UAS-based (DPMaerial) estimation, respectively, to a single, selected, and representative plot correspondent to a specific field issue. The green dotted line means the threshold to determine DPMaerial for GLI index, and the solid black line in each panel indicates the model-fitted values. The black points in each fitted curve

means the actual data correspondent to DPMaerial from the measured plot. Plant maturity estimated by Lodging: 11 DPMaerial and 24 DPMground; Weeds: 26 DPMaerial and 20 DPMground; Low germination: 11 DPMaerial and 21 DPMground; Heterogeneity: 29 DPMaerial and 20 DPMground.

Although these combined issues occurred in just 5% of the entire dataset, outliers could significantly impact the correlation analysis. Correlations were improved after DPMground-related outliers were removed (Supplementary Figure S3 vs. Figure 7). This was especially true for RO in which it increases from $R^2 = 0.57$ to $R^2 = 0.70$ after outlier removal. Additionally, the absolute difference at the plot level compared to DPMground and DPMaerial in most of the cases ranged from 0 to 4 days (Figure 7). Thus, the present pipeline could show a reliable accuracy considering the range of suitable maturity group adaptation zones according to the relative maturity designation and the breeding program strategies. Additionally, in order to comparisons, the correlations, the *RMSE*, and *H* within and across individual trials also was computed and displayed in Supplementary Table 1.

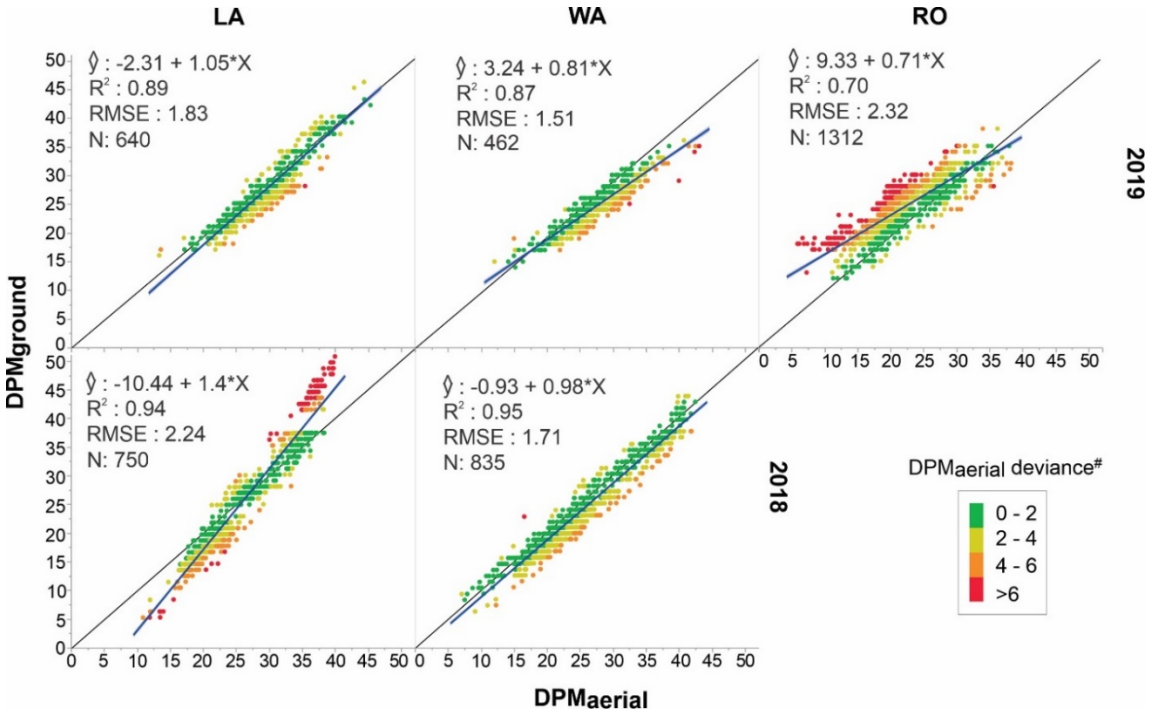


Figure 7. Coefficient of determination (R^2), root mean square deviation (*RMSE*), and number of plots (*N*) for the linear relationship between ground-truth measurements (DPMground) and UAS-based estimated (DPMaerial). Ninety-eight outliers due to DPMground off were removed in total, and 1x/week, GLI index, average pixels values extracted via LOESS model analysis in 2018 and 2019 locations were adopted to perform the analysis. The dark black solid diagonals represent 1:1 line. #Deviance in absolute valor between DPMground and DPMaerial.

Future research can be done to improve plant maturity estimation of soybean research plots. The impact of spatial resolution is one area that can be investigated further. Some studies suggest that high spatial resolution images could be better used for precision agriculture and crop phenotyping applications (Zhang and Kovacs, 2012; Sankaran et al., 2015; Jin et al., 2017). If researchers decide to alter flight altitude or sensors, they must also consider the tradeoffs in cost and time. Due to very reliable results obtained on the current pipeline performing missions with less than 15 minutes by location using a low-cost drone, increasing the GSD may not be justified depending on the propose of the program.

Agronomic crops with senescence prior to harvest also can be studied to apply similar pipeline adopted in this study on soybean for the quantification of maturity, as well as other phenotypes that exhibit similar temporal dependencies. By further, it could be used for understanding the interaction between genetics and the environment at a high-throughput and objective scale for cross-location inferences. As a possible advantage using the SEG model in the future works, the slope value from the senescence curve of the linear regression model can be used to correlate with yield selection in soybeans. Thus, aiming select late or early soybean genotypes, the senescence slope curve has the potential to identified correlated target traits such as yield and performing genotyping recommendation to a specific environment (i.e., higher values of the slope curve aiming to select early soybean lines correlated with gains yield). The slope curve value obtained from the senescence phase also can be used as a trait into a selection index aiming to select soybeans variety according to their growth adaptation (or maturity group).

5. Conclusions

The leading role in this work was to validate diverse strategies to improve the DPMaerial by offering a low-cost pipeline for indicating the proper time of soybean maturity through the greenness quantification. The observed greenness captured using a low-cost UAS-based camera and quantified by the GLI vegetation index fitted via the LOEES regression model and using the mean strategy to pixel extractions method, was able to provide accurate maturity date of soybeans in several environments. Under those conditions, one flight per week was enough to estimate similar or better accurate days of plant maturity compared to DPMground measurements. The underlying reasons for a practical issue in a soybean breeding was assessed checking the significance of deviations between DPMaerial and DPMground. There were found four major issues in this study, dissected in a practical DPMaerial in soybean breeding program

applications: lodging, weeds, low germination, and heterogeneity. Under those practical conditions, the identified outliers decreased considerably the correlation accuracy. Additionally, this feasible phenotyping breeding study offered a convenient package/script plant maturity pipeline to assist breeders and related researchers to be applied in a practical soybean breeding program.

6. Acknowledgments

We acknowledge the United Soybean Board, Minnesota Soybean Research and Promotion Council, and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001 for funding this research. In addition, we thank members of the Lorenz Lab at the University of Minnesota for helping to plant, manage field plots, and take field measurements. Special thanks for due to Thomas Hoverstad, Leonardo Moros, and Steve Quiring for collecting UAS imagery.

7. References

Araus, J.L., S.C. Kefauver, M. Zaman-Allah, M.S. Olsen, and J.E. Cairns. 2018. Translating High-Throughput Phenotyping into Genetic Gain. Trends Plant Sci. 23(5): 451–466. doi: 10.1016/j.tplants.2018.02.001.

Bai, G., Y. Ge, W. Hussain, P.S. Baenziger, and G. Graef. 2016. A multi-sensor system for high throughput field phenotyping in soybean and wheat breeding. Comput. Electron. Agric. 128: 181–192. doi: 10.1016/j.compag.2016.08.021.

Belgiu, M., and A. Stein. 2019. Spatiotemporal image fusion in remote sensing. Remote Sens. 11(7). doi: 10.3390/rs11070818.

Bivand, R., T. Keitt, and B. Rowlingson. 2020. Package ‘ rgdal ’ R. https://cran.r-project.org/web/packages/rgdal/index.html

Castelao Tetila, E., B. Brandoli Machado, N.A.D.S. Belete, D.A. Guimaraes, and H. Pistori. 2017. Identification of Soybean Foliar Diseases Using Unmanned Aerial Vehicle Images. IEEE Geosci. Remote Sens. Lett. 14(12): 2190–2194. doi: 10.1109/LGRS.2017.2743715.

Chapman, S., T. Merz, A. Chan, P. Jackway, S. Hrabar, et al. 2014. Pheno-Copter: A Low-Altitude, Autonomous Remote-Sensing Robotic Helicopter for High-Throughput Field-Based Phenotyping. Agronomy 4(2): 279–301. doi: 10.3390/agronomy4020279.

Chawade, A., J. Van Ham, H. Blomquist, O. Bagge, E. Alexandersson, et al. 2019. High-throughput field-phenotyping tools for plant breeding and precision agriculture. Agronomy 9(5): 1–18. doi: 10.3390/agronomy9050258.

Christenson, B.S., W.T. Schapaugh, N. An, K.P. Price, V. Prasad, et al. 2016. Predicting soybean relative maturity and seed yield using canopy reflectance. Crop Sci. 56(2): 625–

643. doi: 10.2135/cropsci2015.04.0237.

Cleveland, W.S., and S.J. Devlin. 1988. Locally weighted regression: An approach to regression analysis by local fitting. J. Am. Stat. Assoc. 83(403): 596–610. doi: 10.1080/01621459.1988.10478639.

Davies, R.B. 1987. Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika 74(1): 33–43. doi: 10.1093/biomet/74.1.33.

Dobbels, A.A., and A.J. Lorenz. 2019. Soybean iron deficiency chlorosis high throughput phenotyping using an unmanned aircraft system. Plant Methods 15(1): 1–9. doi: 10.1186/s13007-019-0478-9.

van Eeuwijk, F.A., D. Bustos-Korts, E.J. Millet, M.P. Boer, W. Kruijer, et al. 2019. Modelling strategies for assessing and increasing the effectiveness of new phenotyping techniques in plant breeding. Plant Sci. 282(December 2017): 23–39. doi: 10.1016/j.plantsci.2018.06.018.

Escadafal, R., A. Belghith, and H. Ben Moussa. 1994. Indices spectraux pour la dégradation des milieux naturels en Tunisie aride. 6ème Symp. Int. Mesures Physiques et Signatures en Télédétection, ISPRS-CNES. Val d'lsère, France. p. 253–259

Fehr, W.R., and C.E. Caviness. 1977. Stages of soybean development. Iowa Agric. Home Econ. Exp. Satation Publ. (80): 1–12. http://lib.dr.iastate.edu/specialreports/87.

Fox, J., S. Weisberg, and B. Price. 2019. Package 'car.' : 608. https://r-forge.r-project.org/projects/car/, https://CRAN.R-project.org/package=car,%0Ahttps://socserv.socsci.mcmaster.ca/jfox/Books/Companion/index.html.

Furbank, R.T., and M. Tester. 2011. Phenomics - technologies to relieve the phenotyping bottleneck. Trends Plant Sci. 16(12): 635–644. doi: 10.1016/j.tplants.2011.09.005.

Gao, F., M. Anderson, C. Daughtry, and D. Johnson. 2018. Assessing the variability of corn and soybean yields in central Iowa using high spatiotemporal resolution multi-satellite imagery. Remote Sens. 10(9). doi: 10.3390/rs10091489.

Großkinsky, D.K., J. Svensgaard, S. Christensen, and T. Roitsch. 2015. Plant phenomics and the need for physiological phenotyping across scales to narrow the genotype-to-phenotype knowledge gap. J. Exp. Bot. 66(18): 5429–5440. doi: 10.1093/jxb/erv345.

Haghighattalab, A., L. González Pérez, S. Mondal, D. Singh, D. Schinstock, et al. 2016. Application of unmanned aerial systems for high throughput phenotyping of large wheat breeding nurseries. Plant Methods 12(1). doi: 10.1186/s13007-016-0134-6.

Han, L., G. Yang, H. Yang, B. Xu, Z. Li, et al. 2018. Clustering Field-Based Maize Phenotyping of Plant-Height Growth and Canopy Spectral Dynamics Using a UAV Remote-Sensing Approach. Front. Plant Sci. 9. doi: 10.3389/fpls.2018.01638.

Hassan, M.A., M. Yang, L. Fu, A. Rasheed, B. Zheng, et al. 2019. Accuracy assessment of plant height using an unmanned aerial vehicle for quantitative genomic analysis in bread wheat. Plant Methods 15(1). doi: 10.1186/s13007-019-0419-7.

Hijmans, R.J. 2020. Package ' raster ' R. https://cran.r-project.org/web/packages/raster/raster.pdf

Holman, F.H., A.B. Riche, A. Michalski, M. Castle, M.J. Wooster, et al. 2016. High throughput field phenotyping of wheat plant height and growth rate in field plot trials using UAV

based remote sensing. Remote Sens. 8(12). doi: 10.3390/rs8121031.

Huang, Y., K.N. Reddy, R.S. Fletcher, and D. Pennington. 2018. UAV Low-Altitude Remote Sensing for Precision Weed Management. Weed Technol. 32(1): 2–6. doi: 10.1017/wet.2017.89.

Huete, A., K. Didan, T. Miura, E.P. Rodriguez, X. Gao, et al. 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens. Environ. 83(2 Special Issue): 195–213. doi: 10.1080/0965156x.2013.836857.

Hunt, E.R., P.C. Doraiswamy, J.E. McMurtrey, C.S.T. Daughtry, E.M. Perry, et al. 2012. A visible band index for remote sensing leaf chlorophyll content at the Canopy scale. Int. J. Appl. Earth Obs. Geoinf. 21(1): 103–112. doi: 10.1016/j.jag.2012.07.020.

Jarquin, D., R. Howard, A. Xavier, and S. Das Choudhury. 2018. Increasing predictive ability by modeling interactions between environments, genotype and canopy coverage image data for soybeans. Agronomy 8(4). doi: 10.3390/agronomy8040051.

Jin, X., S. Liu, F. Baret, M. Hemerlé, and A. Comar. 2017. Estimates of plant density of wheat crops at emergence from very low altitude UAV imagery. Remote Sens. Environ. 198: 105–114. doi: 10.1016/j.rse.2017.06.007.

Krause, M.R., L. González-Pérez, J. Crossa, P. Pérez-Rodríguez, O. Montesinos-López, et al. 2019. Hyperspectral Reflectance-Derived Relationship Matrices for Genomic Prediction of Grain Yield in Wheat. G3: Genes|Genomes|Genetics: g3.200856.2018. doi: 10.1534/g3.118.200856.

Liaw, A., and M. Wiener. 2018. Package 'randomForest.' https://www.stat.berkeley.edu/~breiman/RandomForests/.

Lindsey, A.J., J.C. Craft, and D.J. Barker. 2020. Modeling canopy senescence to calculate soybean maturity date using NDVI. Crop Sci. 60(1): 172–180. doi: 10.1002/csc2.20079.

Loladze, A., F.A. Rodrigues, F. Toledo, F. San Vicente, B. Gérard, et al. 2019. Application of Remote Sensing for Phenotyping Tar Spot Complex Resistance in Maize. Front. Plant Sci. 10. doi: 10.3389/fpls.2019.00552.

Louhaichi, M., M.M. Borman, and D.E. Johnson. 2001. Spatially located platform and aerial photography for documentation of grazing impacts on wheat. Geocarto Int. 16(1): 65–70. doi: 10.1080/10106040108542184.

Ma, B.L., L.M. Dwyer, C. Costa, E.R. Cober, and M.J. Morrison. 2001. Early prediction of soybean yield from canopy reflectance measurements. Agron. J. 93(6): 1227–1234. doi: 10.2134/agronj2001.1227.

Madec, S., F. Baret, B. de Solan, S. Thomas, D. Dutartre, et al. 2017. High-Throughput Phenotyping of Plant Height: Comparing Unmanned Aerial Vehicles and Ground LiDAR Estimates. Front. Plant Sci. 8. doi: 10.3389/fpls.2017.02002.

Maes, W.H., and K. Steppe. 2019. Perspectives for Remote Sensing with Unmanned Aerial Vehicles in Precision Agriculture. Trends Plant Sci. 24(2): 152–164. doi: 10.1016/j.tplants.2018.11.007.

Mourtzinis, S., and S.P. Conley. 2017. Delineating soybean maturity groups across the United States. Agron. J. 109(4): 1397–1403. doi: 10.2134/agronj2016.10.0581.

Muggeo, V.M.R. 2003. Estimating regression models with unknown break-points. Stat. Med. 22(19): 3055–3071. doi: 10.1002/sim.1545.

Muggeo, V.M.R. 2019. Package 'segmented.' https://cran.r-project.org/web/packages/segmented/segmented.pdf.

Narayanan, B., B. Floyd, K. Tu, L. Ries, and N. Hausmann. 2019. Improving soybean breeding using UAS measurements of physiological maturity. (May): 27. doi: 10.1117/12.2519072.

Peña, J.M., J. Torres-Sánchez, A. Serrano-Pérez, A.I. de Castro, and F. López-Granados. 2015. Quantifying efficacy and limits of unmanned aerial vehicle (UAV) technology for weed seedling detection as affected by sensor resolution. Sensors (Switzerland) 15(3): 5609–5626. doi: 10.3390/s150305609.

Piepho, H.P., and J. Möhring. 2007. Computing heritability and selection response from unbalanced plant breeding trials. Genetics 177(3): 1881–1888. doi: 10.1534/genetics.107.074229.

Reynolds, M., S. Chapman, L. Crespo-Herrera, G. Molero, S. Mondal, et al. 2020. Breeder Friendly Phenotyping. Plant Sci.: 110396. doi: 10.1016/j.plantsci.2019.110396.

Sankaran, S., L.R. Khot, and A.H. Carter. 2015. Field-based crop phenotyping: Multispectral aerial imaging for evaluation of winter wheat emergence and spring stand. Comput. Electron. Agric. 118: 372–379. doi: 10.1016/j.compag.2015.09.001.

Schirrmann, M., A. Hamdorf, A. Garz, A. Ustyuzhanin, and K.H. Dammer. 2016. Estimating wheat biomass by combining image clustering with crop height. Comput. Electron. Agric. 121: 374–384. doi: 10.1016/j.compag.2016.01.007.

Schoving, C., C.O. Stöckle, C. Colombet, L. Champolivier, P. Debaeke, et al. 2020. Combining Simple Phenotyping and Photothermal Algorithm for the Prediction of Soybean Phenology: Application to a Range of Common Cultivars Grown in Europe. Front. Plant Sci. 10(January): 1–14. doi: 10.3389/fpls.2019.01755.

Schweiger, R., S. Kaufman, R. Laaksonen, M.E. Kleber, W. März, et al. 2016. Fast and Accurate Construction of Confidence Intervals for Heritability. Am. J. Hum. Genet. 98(6): 1181–1192. doi: 10.1016/j.ajhg.2016.04.016.

Setiyono, T.D., A. Weiss, J. Specht, A.M. Bastidas, K.G. Cassman, et al. 2007. Understanding and modeling the effect of temperature and daylength on soybean phenology under high-yield conditions. F. Crop. Res. 100(2–3): 257–271. doi: 10.1016/j.fcr.2006.07.011.

Singh, D., X. Wang, U. Kumar, L. Gao, M. Noor, et al. 2019. High-Throughput Phenotyping Enabled Genetic Dissection of Crop Lodging in Wheat. Front. Plant Sci. 10. doi: 10.3389/fpls.2019.00394.

Song, Y., and J. Wang. 2019. Winter Wheat Canopy Height Extraction from UAV-Based Point Cloud Data with a Moving Cuboid Filter. : 10–14.

Strobl, C., A.L. Boulesteix, A. Zeileis, and T. Hothorn. 2007. Bias in random forest variable importance measures: Illustrations, sources and a solution. BMC Bioinformatics 8. doi: 10.1186/1471-2105-8-25.

Thorp, K.R., and L.F. Tian. 2004. A review on remote sensing of weeds in agriculture. Precis. Agric. 5(5): 477–508. doi: 10.1007/s11119-004-5321-1.

Torres-Sánchez, J., F. López-Granados, A.I. de Castro, and J.M. Peña-Barragán. 2013. Configuration and Specifications of an Unmanned Aerial Vehicle (UAV) for Early Site Specific Weed Management. PLoS One 8(3). doi: 10.1371/journal.pone.0058210.

Wang, C., S. Nie, X. Xi, S. Luo, and X. Sun. 2017. Estimating the biomass of maize with

hyperspectral and LiDAR data. Remote Sens. 9(1). doi: 10.3390/rs9010011.

Wolak, M. 2019. Package ‘ nadiv .’

Yang, G., J. Liu, C. Zhao, Z. Li, Y. Huang, et al. 2017. Unmanned Aerial Vehicle Remote Sensing for Field-Based Crop Phenotyping: Current Status and Perspectives. Front. Plant Sci. 8(June). doi: 10.3389/fpls.2017.01111.

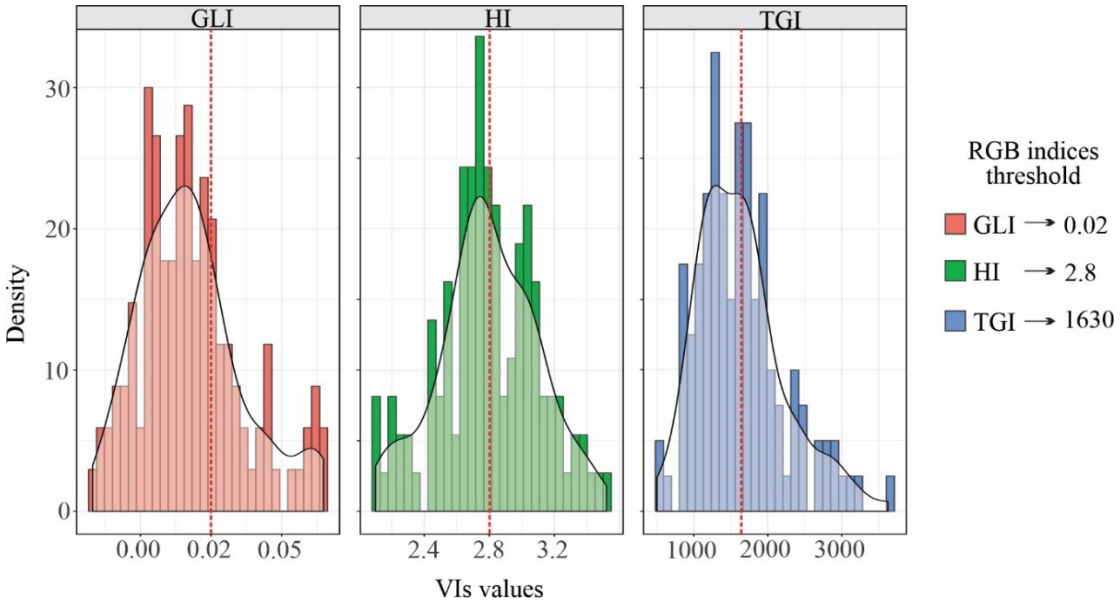
Yu, N., L. Li, N. Schmitz, L.F. Tian, J.A. Greenberg, et al. 2016. Development of methods to improve soybean yield estimation and predict plant maturity with an unmanned aerial vehicle based platform. Remote Sens. Environ. 187: 91–101. doi: 10.1016/j.rse.2016.10.005.

Zdziarski, A.D., M.H. Todeschini, A.S. Milioli, L.G. Woyann, A. Madureira, et al. 2018. Key soybean maturity groups to increase grain yield in Brazil. Crop Sci. 58(3): 1155–1165. doi: 10.2135/cropsci2017.09.0581.

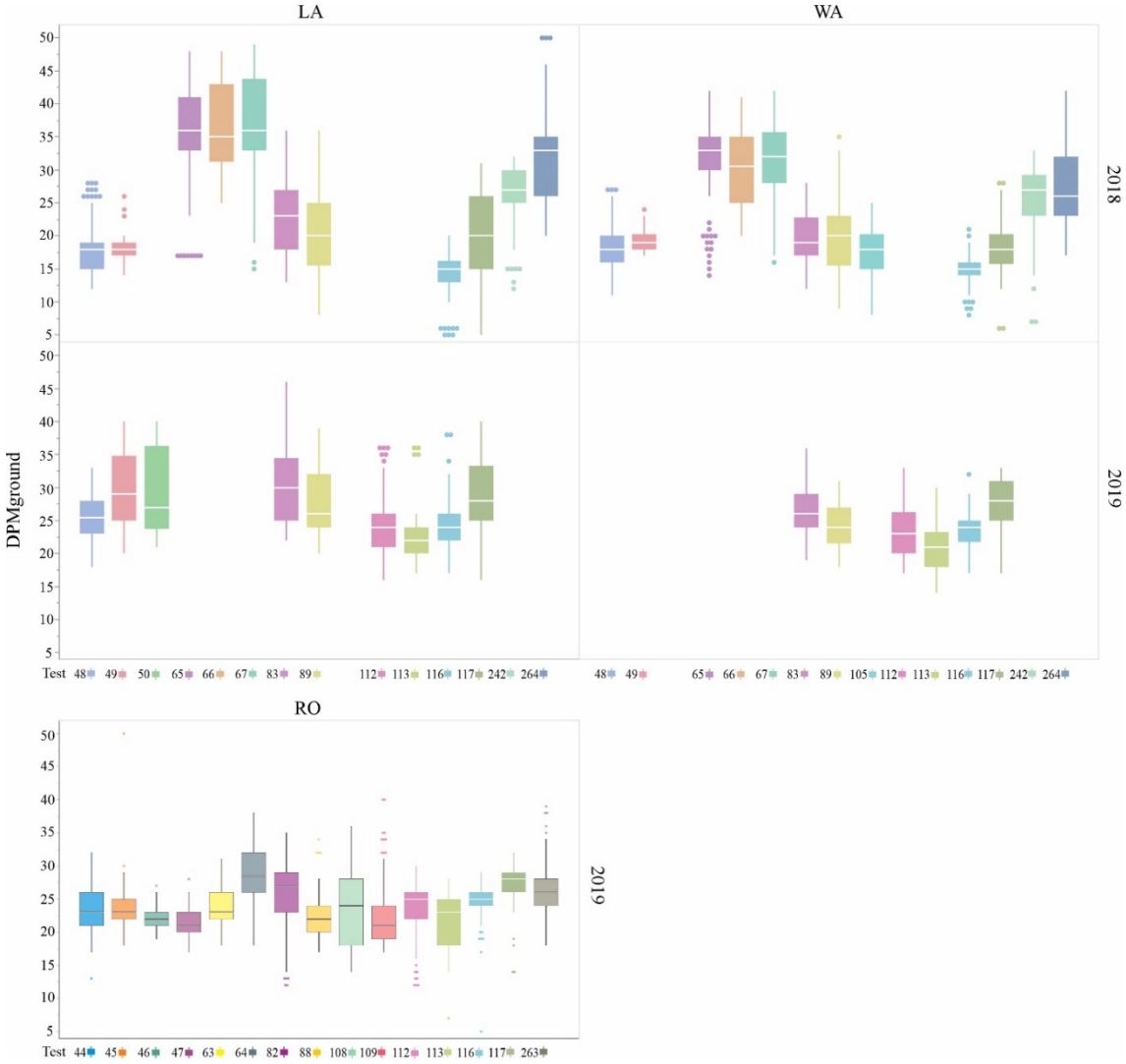
Zhang, C., and J.M. Kovacs. 2012. The application of small unmanned aerial systems for precision agriculture: A review. Precis. Agric. 13(6): 693–712. doi: 10.1007/s11119-012-9274-5.

Zhou, J., D. Yungbluth, C.N. Vong, A. Scaboo, and J. Zhou. 2019. Estimation of the maturity date of soybean breeding lines using UAV-based multispectral imagery. Remote Sens. 11(18). doi: 10.3390/rs11182075.

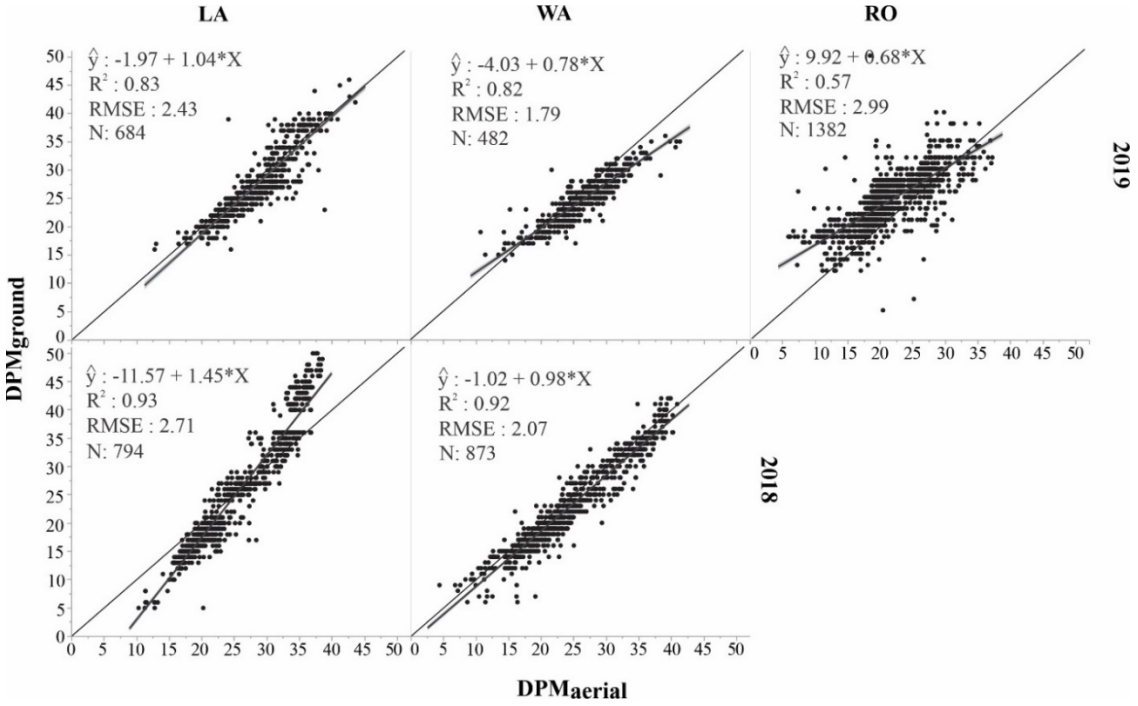
8. Supplemental Material



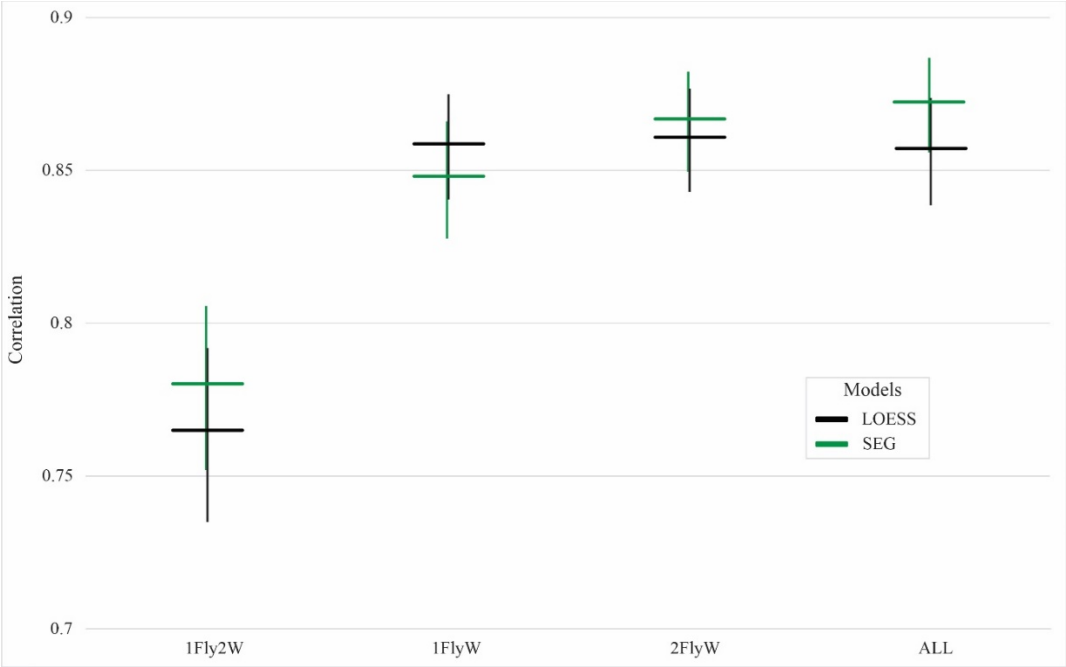
Supplementary Fig. S1. Distribution of vegetation index (VI) value of soybean plots that were scored as mature using ground-based visual ratings. These plots matured on the same day as a corresponding UAS flight. GLI: greenness leaf index; HI: Hue index; TGI: triangular greenness index. The dotted red vertical lines indicate the threshold selected according to the optimized correlated VIs values.



Supplementary Fig. S2. Boxplots of ground-truth measurements of plant maturity date (DPMground). Data is separated by environment and trials. Tests are given arbitrary numbers based on their stage in the cultivar development pipeline. LA: Lamberton; WA: Waseca; RO: Rosemount locations.



Supplementary Fig. S3. Coefficient of determination (R^2), root mean square deviation (*RMSE*), and number of plots (*N*) for the linear relationship between ground-truth measurements (DPMground) and UAS-based estimated (DPMaerial). One flight per week, GLI VI, and mean pixels values extracted were adopted to perform the correlations analysis via the LOESS model in 2018 and 2019 locations. The dark black solid diagonals represent 1:1 line. LA: Lamberton, WA: Waseca, and RO: Rosemount locations.



Supplementary Fig. S4. Correlation between ground-truth (DPMground) and UAS-based (DPMaerial) plant maturity estimation using four different timing of flights under Local (LOESS) and Segmented (SEG) regression models across 2019 locations. Settings of flights performed 1Fly2W: one flight every two weeks; 1FlyW: average value from the one flight per week settings; 2Fly_W: two flights per week; ALL: all flights available.

Supplementary Table S1. Correlations (*r*), root mean square deviation (*RMSE*) and broad-sense heritability (*H*) for 53 trials in five locations contrasting UAS-based (DPMaerial) estimates and ground-truth measurements (DPMground) via LOESS model using GLI index, average pixels values extracted, and one flight per week (1x/week).

				Corr. (*r*)[#]		*RMSE*		*H ground*	*H UAS*	*H ground*	*H UAS*
Year	**Site[£]**	**TN***	**NP****	**+**	**-**	**+**	**-**	**+**	**+**	**-**	**-**
2018	LA	48	94	0.85	**0.89**	2.10	**1.82**	**0.87**	0.86	**0.93**	0.86
2018	LA	49	62	0.64	**0.71**	1.59	**1.44**	0.65	**0.73**	0.71	**0.84**
2018	LA	65	95	0.92	**0.92**	3.11	**2.78**	0.94	**0.94**	0.93	**0.94**
2018	LA	66	64	0.93	**0.95**	2.57	**2.10**	0.92	**0.94**	**0.96**	0.94
2018	LA	67	64	0.94	**0.96**	2.81	**2.23**	0.93	**0.96**	**0.97**	0.94
2018	LA	83	117	0.94	**0.95**	1.99	**1.94**	0.93	**0.95**	0.94	**0.95**
2018	LA	89	33	0.94	**0.95**	2.47	**2.28**	**1.00**	0.96	**1.00**	0.96
2018	LA	116	94	0.93	**0.93**	1.27	**1.27**	0.75	**0.77**	0.75	**0.77**
2018	LA	117	62	0.90	**0.95**	2.52	**1.72**	0.82	**0.92**	**0.93**	0.93
2018	LA	242	62	0.94	**0.95**	1.83	**1.57**	0.94	**0.94**	0.93	**0.96**
2018	LA	264	47	0.92	**0.94**	3.11	**2.12**	**0.98**	0.90	**0.98**	0.87
2018	WA	48	94	0.86	**0.91**	1.80	**1.46**	0.73	**0.76**	0.74	**0.78**
2018	WA	49	62	0.82	**0.82**	1.07	**1.07**	0.40	**0.44**	0.40	**0.44**
2018	WA	65	92	0.94	**0.97**	2.16	**1.65**	0.86	**0.91**	0.93	**0.94**
2018	WA	66	63	0.95	**0.96**	1.85	**1.59**	0.88	**0.90**	0.88	**0.90**
2018	WA	67	63	0.95	**0.96**	2.07	**1.84**	**0.94**	0.94	**0.96**	0.93
2018	WA	83	116	0.92	**0.92**	1.29	**1.26**	0.69	**0.77**	0.67	**0.76**
2018	WA	89	33	0.97	**0.98**	1.55	**1.38**	0.95	**0.96**	0.95	**0.97**
2018	WA	105	78	0.94	**0.95**	1.32	**1.15**	0.92	**0.92**	0.92	**0.94**
2018	WA	116	94	0.89	**0.89**	1.11	**1.11**	0.56	**0.74**	0.56	**0.74**
2018	WA	117	62	0.93	**0.94**	1.62	**1.45**	**0.92**	0.91	0.91	**0.94**
2018	WA	242	62	0.93	**0.94**	2.80	**1.85**	**0.95**	0.90	0.95	**0.96**
2018	WA	264	47	0.91	**0.94**	2.62	**2.05**	**0.76**	0.58	0.87	**0.78**
2019	LA	48	62	0.88	**0.92**	1.71	**1.41**	**0.82**	0.79	**0.83**	0.78
2019	LA	49	62	0.90	**0.95**	2.65	**1.81**	**0.91**	0.80	**0.90**	0.86
2019	LA	50	78	0.95	**0.96**	1.94	**1.69**	**0.91**	0.85	**0.90**	0.86
2019	LA	83	117	0.93	**0.95**	2.23	**1.89**	**0.32**	0.25	**0.32**	0.25
2019	LA	89	69	0.88	**0.92**	2.53	**1.93**	**0.86**	0.82	0.86	**0.86**

2019	LA	112	78	0.94	**0.94**	1.57	**1.57**	**0.92**	0.89	**0.92**	0.89
2019	LA	113	62	0.91	**0.93**	1.64	**1.23**	**0.97**	0.86	**0.95**	0.83
2019	LA	116	94	0.91	**0.91**	1.57	**1.33**	**0.87**	0.70	**0.82**	0.65
2019	LA	117	62	0.74	**0.92**	3.87	**2.02**	0.61	**0.89**	**0.92**	0.90
2019	RO	44	62	0.78	**0.83**	2.71	**1.88**	0.79	**0.91**	0.76	**0.93**
2019	RO	45	78	0.36	**0.56**	3.76	**2.10**	0.44	**0.67**	0.44	**0.72**
2019	RO	46	62	0.74	**0.74**	1.22	**1.22**	0.77	**0.87**	0.77	**0.87**
2019	RO	47	62	0.78	**0.78**	1.47	**1.48**	**0.80**	0.80	0.80	**0.80**
2019	RO	63	72	0.82	**0.84**	1.71	**1.63**	0.74	**0.82**	0.80	**0.82**
2019	RO	64	144	0.70	**0.81**	3.09	**2.31**	0.63	**0.84**	0.78	**0.86**
2019	RO	82	147	0.87	**0.87**	2.66	**2.66**	0.92	**0.94**	0.95	**0.95**
2019	RO	88	135	0.84	**0.85**	1.80	**1.72**	**0.85**	0.84	**0.89**	0.84
2019	RO	108	67	0.75	**0.83**	3.43	**2.89**	0.76	**0.80**	**0.85**	0.82
2019	RO	109	141	0.66	**0.83**	3.40	**2.10**	0.56	**0.87**	0.74	**0.89**
2019	RO	112	78	0.80	**0.85**	2.66	**2.11**	**0.95**	0.77	**0.94**	0.83
2019	RO	113	62	0.53	**0.77**	3.70	**2.49**	0.53	**0.84**	**0.88**	0.83
2019	RO	116	94	0.53	**0.68**	2.62	**1.71**	0.36	**0.72**	0.71	**0.71**
2019	RO	117	62	0.68	**0.78**	2.59	**1.95**	**0.94**	0.71	**0.94**	0.83
2019	RO	263	111	0.67	**0.83**	2.52	**1.69**	**0.86**	0.77	**0.87**	0.75
2019	WA	83	117	0.93	**0.94**	1.42	**1.33**	**0.75**	0.60	**0.75**	0.60
2019	WA	89	69	0.85	**0.90**	1.71	**1.40**	**0.78**	0.70	**0.76**	0.73
2019	WA	112	78	0.89	**0.93**	1.94	**1.53**	**0.92**	0.89	0.92	**0.93**
2019	WA	113	62	0.86	**0.93**	2.04	**1.42**	0.81	**0.83**	**0.90**	0.82
2019	WA	116	94	0.90	**0.91**	1.30	**1.24**	0.72	**0.82**	0.75	**0.82**
2019	WA	117	62	0.91	**0.93**	1.74	**1.48**	**0.93**	0.86	0.93	**0.95**

[#]Correlations (*r*) between DPMaerial and DPMground; *Trial number (TN); **number of plots (NP). Columns with (+) and (-) are the dataset with and without identified outliers, respectively; [£]Locations, LA: Lamberton, RO: Rosemount, and WA: Waseca

GENERAL CONCLUSIONS

This work's leading role was to implement and validate feasible and scalable remote sensing techniques (HTP via UAS-based) to acquire PHs and DPM insights. These phenological traits provide valuable data to wheat and soybean breeders.

The developed semi-automatic workflow through remote sensing techniques allowed to access PHs obtained from CSM. These results were obtained using a UAS-based platform produced by 3D topographic reconstructions of the crop fields with and without vegetation. The UAS-PPK/RTK approach showed a reliable method to replace the laborious conventional methods using GCPs. Our findings also indicate that using different UAS-platforms for phenotyping selection could be cost-effective to estimated PHs values, mainly in late GS using UAS-based.

The estimation of DPM demonstrated a low-cost pipeline for indicating the proper time of soybean maturity and provided an interactive open-source and efficient method to assist breeders and related scientists in practical conditions. Thus, the soybean maturity date was successfully estimated using drone imagery by modeling the decay of canopy greenness through time. However, plant lodging, presence of weeds, low germination, and within-line genetic heterogeneity should be taken into consideration to inaccurate UAS-based maturity date estimates. Crops including dry beans, canola, sorghum, corn, rice, wheat, sunflower, oats, and millet all go through senescence prior to harvest, hence similar pipelines adopted in this study on soybean can be applied for the quantification of maturity in these crops, as well as, other phenotypes that exhibit similar temporal dependencies.