University of São Paulo
"Luiz de Queiroz" College of Agriculture

High-throughput phenotyping via UAS: the optimization within a breeding program and a new validation method based on simulation

Giovanni Galli

Thesis presented to obtain the degree of Doctor in Science. Area: Genetics and Plant Breeding

Piracicaba
2020

Giovanni Galli
Agronomist

High-throughput phenotyping via UAS: the optimization within a breeding program and a new validation method based on simulation

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Advisor:
Prof. Dr. **ROBERTO FRITSCHE NETO**

Thesis presented to obtain the degree of Doctor in Science. Area: Genetics and Plant Breeding

Piracicaba
2020

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA – DIBD/ESALQ/USP

Galli, Giovanni

High-throughput phenotyping via UAS: the optimization within a breeding program and a new validation method based on simulation / Giovanni Galli. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2020.

71 p.

Tese (Doutorado) - - USP / Escola Superior de Agricultura “Luiz de Queiroz”.

1. Fenômica 2. In silico 3. Imageamento aéreo 4. Sorgo 5. Structure-from-motion I. Título

DEDICATÓRIA

À minha família.

Dedico

AGRADECIMENTOS

Àos meus pais e irmãos pelo apoio e amor incondicional.

À minha esposa, Cassia, pela compreensão, incentivo e carinho.

Ao Professor Dr. Roberto Fritsche Neto pela oportunidade, ensinamentos, orientação e exemplo profissional.

Aos colegas do Laboratório de Melhoramento de Plantas Alógamas pelo auxílio sempre que necessário e pela amizade que levarei para toda a vida.

À Texas A&M University, ao Dr. William L. Rooney e ao Sorghum Breeding Lab pela receptividade.

À CAPES (código de financiamento 001) e ao CNPq pela concessão da bolsa doutorado.

Àqueles aqui não mencionados que de alguma forma contribuíram para a realização deste trabalho.

CONTENTS

RESUMO

Fenotipagem de alto rendimento via VANTs: a otimização em um programa de melhoramento e um novo método de validação baseado em simulação

A Fenotipagem de alto rendimento (HTP), ou simplesmente fenômica, tem chamado a atenção da comunidade científica como uma área com potencial de aumentar a custo-efetividade e acurácia de fenotipagem. Entretanto, a viabilidade deste conjunto de abordagens ainda precisa confirmação. Neste contexto, dois grandes desafios para a seu emprego são a otimização do uso de dados (*data-to-decision*) e a validação de procedimentos para cenários específicos de seleção. Nós acrescentamos a este tema reportando resultados de dois estudos que objetivaram a otimização e validação de HTP para experimentos a campo baseada em veículos aéreos não tripulados (VANTs). No primeiro, apresentamos uma prova de conceito usando dados de sorgo granífero com o objetivo de identificar quando os dados de HTP devem ser coletados e como devem ser processados para a otimização da predição de dois caracteres de importância agronômica, produtividade de grãos e sanidade de planta. Nossos resultados sugerem que não há incremento da capacidade preditiva quando múltiplos índices vegetativos e voos são combinados. Adicionalmente, um único índice e voo pode ser usado para predizer ambas características sem perda expressiva de acurácia. No segundo, apresentamos uma nova ferramenta para validação de abordagens de HTP baseadas em imagens aéreas com uso de simulações de computador. A ferramenta foi exemplificada com um estudo de caso de mensuração de altura de plantas em milho. Nossos resultados sugerem que os experimentos gerados *in silico* podem ser adequadamente reconstruídos com algoritmos de *structure-from-motion* usando imagens renderizadas, permitindo a realização de inferências sobre os fatores testados. Este estudo também trouxe novos conhecimentos sobre o efeito de fatores experimentais sobre a acurácia da mensuração de altura de plantas usando HTP. Por fim, acreditamos que nossos resultados permitirão a compreensão mais profunda da prática da HTP, auxiliando os melhoristas na busca por uma seleção mais confiável e custo-efetiva.

Palavras-chave: Fenômica, In silico, Imageamento aéreo, Sorgo, Structure-from-motion

ABSTRACT

High-throughput phenotyping via UAS: the optimization within a breeding program and a new validation method based on simulation

High-throughput phenotyping (HTP), or simply phenomics, has drawn the attention of the scientific community as a field with the potential to increase phenotyping cost-effectiveness and accuracy. Nevertheless, the feasibility of this set of approaches is yet to be confirmed. In this sense, two major challenges to its application are optimizing the data-to-decision process and the validation of procedures and pipelines for specific selection scenarios. We add to this matter by reporting on two studies aimed at the optimization and validation of field HTP based on unmanned aerial systems (UAS). In the first, we presented a proof-of-concept investigation using a grain sorghum dataset with the intent of identifying when HTP data should be collected and how it should be processed for the optimization of prediction of two major traits, grain yield and plant health. Our findings suggest that there is no predictive ability increase when combining multiple vegetation indices and flight dates. Additionally, a single index and flight can be used for predicting both traits without expressive accuracy loss. In the second, we presented a new tool for validating aerial image-based HTP approaches with computer simulations. The approach was exemplified with a comprehensive study case of plant height estimation in maize. Our results show that the *in silico* experiments could be adequately reconstructed with structure-from-motion algorithms using UAS-like rendered images, enabling inference-making about tested factors. This study also brought new insights into the effect of experimental factors over the accuracy of plant height assessment using HTP. At last, we believe that our findings allowed the promotion of a deeper understanding of the HTP practice, enabling breeders to work towards a more reliable and cost-effective selection.

Keywords: Phenomics, In silico, Aerial imagery, Sorghum, Structure-from-motion

1. INTRODUCTION

Genetic gain is the rate of performance increase achieved by artificial selection per unit time. This key concept of plant breeding is defined as a function of selection intensity, selection accuracy, genetic variance, and time per selection cycle (Araus et al., 2018). High-throughput phenotyping (HTP), also known as phenomics, has taken the spot as a field with potential to act over these parameters, enabling breeding programs to increase the rates of gain (White et al., 2012; Araus and Cairns, 2014; Araus et al., 2018). Among the realm of available HTP devices, special attention has been paid to unmanned aerial systems (UAS) due to its relatively low cost and ability to carry out large-scale screenings at high temporal/spatial resolution. Nevertheless, HTP effectiveness still needs proof. Therefore, new approaches, pipeline optimizations, and proof-of-concept studies are highly valuable for the improvement and feasibility assessment of UAS-based HTP, producing results of immediate and direct benefit to the scientific community.

One of the major challenges for the effective application of HTP is overcoming the data-to-decision barrier. Recent technological advancements have made data collection no longer a bottleneck as the high temporal/spatial resolution of HTP is capable of generating vast amounts of data (Araus and Kefauver, 2018). Nevertheless, given that the set of phenotypes of an organism is highly conditional and dynamic across its life cycle (Cobb et al., 2013), identifying meaningful genetic signals from data and efficiently translating the information to target traits is key (Araus and Kefauver, 2018). One approach towards optimized phenotyping is by finding the most informative phenological stage(s) and high throughput traits for a given selection purpose. This could work towards overcoming the technical data-to-decision difficulty enabling the application of truly high-throughput phenotyping.

Another key challenge to the application of HTP in plant breeding is the validation of procedures and pipelines for specific selection scenarios. In this sense, controlled and representative experimental conditions are required for meaningful and generalist inferencing. However, these conditions are sometimes not easy to achieve. A usual way to bypass the limitations of real datasets is by implementing simulations (Barth et al., 2018; Müller et al., 2018). This could provide the process of validating HTP methodologies with adequate conditions, potentially improving the reliability of estimates. Additionally, by employing simulations into the HTP context, one would be able to test new methods, experimental setups, measure the potential of a given approach, and dimension equipment, all of this without performing any field experiments or acquiring phenotyping gear. However, to this date, no report on the use of simulations for HTP validation has been made.

In the following sections, we report on two studies aimed at validating and optimizing HTP of field experiments. In the first, we presented a research as to when UAS imagery data should be collected and how it should be processed for the optimization of the phenotyping for grain yield and overall plant health (based on anthracnose scores) in grain sorghum. In the second, we proposed a novel way of designing and validating HTP approaches based on aerial imagery with computer-generated tridimensional experiments. This new tool is exemplified with a comprehensive case study of plant height estimation in maize under contrasting experimental setups in the context of plant breeding.

REFERENCES

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

Araus, J.L., and S.C. Kefauver. 2018. Breeding to adapt agriculture to climate change: affordable phenotyping solutions. Curr. Opin. Plant Biol. doi: 10.1016/j.pbi.2018.05.003.

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.

Barth, R., J. IJsselmuiden, J. Hemming, and E.J. Van Henten. 2018. Data synthesis methods for semantic segmentation in agriculture: A Capsicum annuum dataset. Comput. Electron. Agric. 144(December 2017): 284–296. doi: 10.1016/j.compag.2017.12.001.

Cobb, J.N., G. DeClerck, A. Greenberg, R. Clark, and S. McCouch. 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(4): 867–887. doi: 10.1007/s00122-013-2066-0.

Müller, M., V. Casser, J. Lahoud, N. Smith, and B. Ghanem. 2018. Sim4CV: A Photo-Realistic Simulator for Computer Vision Applications. Int. J. Comput. Vis. 126(9): 902–919. doi: 10.1007/s11263-018-1073-7.

White, J.W., P. Andrade-Sanchez, M.A. Gore, K.F. Bronson, T.A. Coffelt, et al. 2012. Field-based phenomics for plant genetics research. F. Crop. Res. 133: 101–112. doi: 10.1016/j.fcr.2012.04.003.

2. OPTIMIZATION OF UAS-BASED HIGH-THROUGHPUT PHENOTYPING FOR TRAITS WITH DIFFERENT GENETIC ARCHITECTURE: ESTIMATION OF PLANT HEALTH AND GRAIN YIELD IN SORGHUM HYBRIDS

ABSTRACT

High-throughput phenotyping (HTP) has enabled the acquisition of vast amounts of data. In this sense, finding the most informative phenological stage(s) and high throughput traits could lead to significant optimization of HTP-assisted selection. Herein, an investigation as to when phenotypic data should be collected and how it should be processed from UAS imagery for the optimization and assessment of two primary traits in grain sorghum, namely, grain yield and overall plant health (based on anthracnose scores) was conducted. By evaluating multiple flight dates across the growing season (69, 91, 102, 119, and 127 days after sowing) via multispectral UAS-based imagery, a set of scenarios composed by combinations of flight dates and vegetation indices (CIg, CIre, NDVI, NDVIre, MSR, and MSRre) were constructed for analysis. We present an exploratory genetic analysis from HTP aerial imagery based on parameters such as repeatability, experimental accuracy, correlation to primary traits (grain yield and overall plant health), and indirect selection efficiency for comparative purposes. A comprehensive focus on the ability to predict primary traits using vegetation indices under each scenario with partial least squares regression and neural networks was practiced. In this sense, results showed no increase in predictive ability when combining multiple vegetation indices. Therefore, it is our assessment that using only an index with a higher predictive ability (e.g., NDVI or MSR for plant health with 0.75; and any index but CIg for grain yield with ~0.55) is recommended. Likewise, the combining of multiple flights did not result in a significant increase in prediction ability for both primary traits. Thus, we observed that a single flight for each trait (e.g., 119 days after sowing with 0.81 for plant health; 102 days after sowing with 0.59 for grain yield) was optimal. Concercing the predictive algorithms examined, partial least squares regression (PLSR) and neural network, results were similar with PLSR generally outperforming. In addition, we discuss our findings from an application standpoint of a field-based breeding program and suggest additional optimization options.

Keywords: neural networks; multispectral; aerial imagery; partial least squares regression; vegetation indices.

2.1. INTRODUCTION

Given the advancements in plant genotyping during the last decade, plant phenotyping is now the limiting factor in the ability to determine the genetic control of important phenotypic traits (Araus and Cairns, 2014). This fact has generated interest to determine how compensation for the lack of high-quality phenotypic information can be obtained (Cobb et al., 2013). High-throughput phenotyping (HTP) or simply, phenomics, has thus been deployed to measure features of plants (Araus and Cairns, 2014).

Vegetation indices (band ratios) collected with HTP can be used as secondary traits for indirect selection of main/primary traits (e.g. grain yield; GY). This indirect selection is based on specific features (e.g. biomass via normalized difference vegetation index (NDVI); chlorophyll content via a chlorophyll index (CI)) that compose agronomically relevant traits (Li and Shi., 2018). In sorghum (*Sorghum bicolor* L. Moench), multiple indices have been shown to be correlated with overall plant health (Stanton et al., 2017; Pugh et al., 2018), GY (Stanton et al., 2017; Shafian et al., 2018; Pugh et al., 2018), leaf area index (LAI) (Shafian et al., 2018), senescence (Potgieter et al., 2017),

fractional vegetation cover (Shafian et al., 2018), biomass yield, and chlorophyll and nitrogen content (Li and Shi, 2018).

The response of an organism to a set of multidimensional exogenous and endogenous signals that occur in an integrated way throughout its development is termed the phenome (Cobb et al., 2013), which is highly conditional and dynamic. Consequently, the association of primary and secondary traits throughout the life cycle of a plant depends on a series of predictable and unpredictable events. Gene expression is temporal; canopy features and their genetic control change over the crop development (Feng et al., 2017). Furthermore, singular events such as droughts can also change this relationship (Shi et al., 2017). Both predictable and unpredictable events condition traits that can be captured by HTP platforms with high temporal resolution. However, with higher temporal resolution, there is an expected increase in redundancy of information along with the need for better storage and processing equipment to handle the astonishing amount of newly generated data. This raises questions on the usefulness of multiple measures based on the complementarity of secondary traits at multiple phenological stages and its implications to understanding or predicting primary complex traits (Araus and Kefauver, 2018).

Both GY and anthracnose (*Colletotrichum sublineola*) resistance are two complex traits of high importance in sorghum breeding. Anthracnose is a fungal disease that drastically affects plant health and limits yield potential in specific environments (da Costa et al., 2011), with resistance being oligogenic in nature (da Costa et al., 2011; Patil et al., 2017). GY is dependent upon various morphological characteristics and is polygenic in nature (Boyles et al., 2017). As these traits are correlated (Pugh et al., 2018), an identification of the appropriate phenological stage(s) at which the signal of genotype effect is maximized for both would lead to significant optimization improvement of HTP-based selection. The relevant cost of phenotyping for both traits make it suitable material for proof-of-concept studies using HTP, producing results of benefit to breeders in the decision-making process.

The aims of this study were to: i) determine the optimal number of evaluation dates needed to predict anthracnose resistance scores (overall plant health) and GY in sorghum hybrids; ii) assess the most appropriate (combination of) vegetation indices based on multispectral imagery in the prediction of anthracnose scores (overall plant health) and GY in sorghum hybrids; iii) compare algorithms for prediction of primary traits (overall plant health and GY); and iv) verify the utility of these tools for sorghum breeding programs.

2.2. MATERIAL AND METHODS

2.2.1. Germplasm and experimental design

A set of 32 experimental and open-pedigree sorghum hybrids were evaluated during the summer of 2018 in College Station, TX [30°32'33.4"N 96°25'46.6"W] using a randomized complete block design with three replications. Experimental plots were sown on April 3rd (0 DAS) and consisted of two 5.5 m plots with a row spacing of 0.762 m, an approximate sowing density of 200,000 plants per hectare. Fungal inoculations of anthracnose were administered 45 DAS by placing infested sorghum seeds within the plant whorl. Inoculate was prepared according to the protocol presented in Pugh et al. (2018). Standard agronomic practices for grain sorghum were used for this study. Traits evaluated in this experiment hereafter are referred to as primary (ground truth/target) and secondary (surrogate/remotely acquired).

2.2.2. Phenotyping of primary traits

Anthracnose severity was measured using the area under disease progress curve (AUDPC). To create the AUDPC, anthracnose incidence and severity scores (0-9) were recorded as the proportion of plants in a given plot showing typical symptoms of anthracnose infection, i.e. grey diseased tissue and or necrotic lesions (Pugh et al., 2018). The utilized scale was adapted from the original to include level 0, which stands for no symptom. Anthracnose measurements were recorded on 11 different dates throughout the growing season (57, 65, 70, 77, 83, 91, 99, 104, 112, 119 and 127 DAS). Anthracnose incidence and severity was then utilized to estimate an anthracnose AUDPC score for plot i by using the equation:

$$AUDPC_i = \sum_{k=1}^{n-1} \left(\frac{y_k + y_{k+1}}{2} \right) (t_{k+1} - t_k)$$

where n is the number of evaluations, k is the time-point, y is the score (disease intensity or incidence), and t is the time (in days) at which the score was taken. AUDPC metrics follow anthracnose scores, meaning that higher values are associated with lower disease resistance. At the end of the field season when grain moisture content was below 15%, the GY was measured by mechanically harvesting plots. For reporting purposes, grain yield (Mg ha^{-1}) of all entries was adjusted to 13% moisture.

2.2.3. Phenotyping of secondary traits (UAS)

Image acquisition for vegetation index estimation was performed using small unmanned aerial systems (UAS) equipped with a MicaSense® RedEdge (MicaSense, Seatle, USA) multispectral camera (blue: 375 nm; green: 560 nm; red: 668 nm; red edge: 717 nm; near infra-red: 840 nm) and/or a SlantRange® 3P (SlantRange, San Diego, USA) multispectral camera (green: 550 nm; red: 650 nm; red edge: 720 nm; near infra-red: 850 nm). As different cameras were utilized, flight altitudes were adjusted to obtain similar ground resolutions. Data captured by UAS were collected across five separate dates beginning at 69 DAS and ending on 127 DAS (Table 1). Images were collected plus or minus 1 to 2 hours of solar noon, minimizing inconsistencies due to solar angle and light scattering, and were subsequently radiometrically calibrated for comparison across multiple flight dates. Flights were projected to achieve image overlapping of 75~80%. Additionally, ground control points (GCPs) were placed around and at the middle of the study site and coordinates recorded using Post-Processing Kinematic (PPK) Global Positioning System (GPS) for precision georeferencing. Additional information regarding data collection is presented in Table 1.

Data acquisition and mosaicking were performed by the Unmanned Aerial System (UAS) Remote Sensing, Digital Agriculture Program at the Texas A&M AgriLife Corpus Christi, TX research and extension center. Image processing was performed in Agisoft Photoscan Pro (Agisoft LLC, St. Petersburg, Russia) version 1.4.3 (build 6529). Processing was comprised of alignment and optimization of image position, generation of point clouds, and generation of digital surface models (DSM) and orthomosaics. Mosaics were imported into QGIS software (version 3.02 Girona; QGIS Development Team, 2018) where plot boundaries were drawn using *v.mkgrid*, a function of GRASS plugin (version 7.4.0; GRASS Development Team, 2018). A negative buffer of 0.15 m was drawn to minimize neighboring effects of plot leaf overlap using the internal *Buffer* function. Orthomosaics and plot shapefiles were loaded and processed in R environment (R Core Team, 2011) using *raster* (Hijmans et al., 2015) and *sp*

(Pebesma and Bivand, 2005) packages to extract relevant vegetation indices (Table 2). The index value is represented by the median of all pixel values in a given plot.

2.2.4. Phenotypic analysis

Primary and secondary traits were subjected to individual analysis using linear mixed modeling for the estimation of variance components and genotypic adjusted means. Adjusted means were estimated using ordinary least squares (OLS), while variance components were estimated using restricted maximum likelihood (REML)/BLUP (Henderson, 1975) with the ASReml-R package (version 3.0; Butler et al., 2007). The model structure was as follows:

$$\boldsymbol{y} = X\boldsymbol{\beta} + H\boldsymbol{r} + V\boldsymbol{c} + T\boldsymbol{g} + \boldsymbol{\varepsilon}$$

where $\boldsymbol{y}$ is the vector of phenotypic observations (primary or secondary traits); $\boldsymbol{\beta}$: is the vector of the fixed effects of replication added to the overall mean; for the analysis of secondary traits, the effect of flight date was also added as fixed. $\boldsymbol{r}$: is the vector of row within replication and is regarded as random [$\boldsymbol{r} \sim N(0, I\sigma_r^2)$]; $\boldsymbol{c}$: is the random effect of column [$\boldsymbol{c} \sim N(0, I\sigma_c^2)$]; $\boldsymbol{g}$: is the vector of the adjusted means for hybrids; and $\boldsymbol{\varepsilon}$: is the vector for error [$\boldsymbol{\varepsilon} \sim N(0, I\sigma_\varepsilon^2)$]. X, H, V, and T are the incidence matrices that relate the independent vectors to the response variable $\boldsymbol{y}$. Wald's and Likelihood Ratio Test (LRT) were utilized for the assessment of significance of fixed and random factors, respectively. To obtain genotypic variance components, the previous model was ran treating $\boldsymbol{g}$ as a random effect [$\boldsymbol{g} \sim N\left(0, I\sigma_g^2\right)$].

Genotypic correlations (r_{ps}) between primary and secondary traits were estimated based on covariances obtained by fitting bivariate models following the same factor structure as the aforementioned model. Both treatment and residual variances from the single-trait analysis were fed as initial values to facilitate convergence. Once covariances were obtained, genotypic correlations were estimated using the following equation:

$$r_{ps} = \frac{cov_{ps}}{\sqrt{\sigma_{gp}^2} . \sqrt{\sigma_{gs}^2}}$$

where cov_{ps} is the covariance between primary and secondary trait; σ_{gp}^2 is the genotypic variance of the primary trait; and σ_{gs}^2 is the genotypic variance of the secondary trait.

2.2.5. Phenotyping quality and indirect selection efficiency

To assess overall data quality, both repeatability and experimental accuracy estimates for primary and secondary traits were estimated. Repeatability (R^2) estimates were conducted at the plot level and on an entry mean level as follows; plot: $R_i^2 = \sigma_g^2/\left(\sigma_g^2 + \sigma_\varepsilon^2\right)$; entry mean: $R_m^2 = \sigma_g^2/\left(\sigma_g^2 + \sigma_\varepsilon^2/r\right)$; where r is the number of replicates. Experimental accuracy was estimated as $r_{gg} = \left[1 - \frac{1}{1+r.CV_R^2}\right]^{1/2}$ given that $CV_R = CV_g/CV_\varepsilon$, considering $CV_g = \left(\sqrt{\sigma_g^2}/\mu\right) \times 100$ and $CV_\varepsilon = \left(\sqrt{\sigma_\varepsilon^2}/\mu\right) \times 100$; where μ is the grand mean of the trait (Resende and Duarte, 2007).

Repeatabilities and genotypic correlations were used to estimate the efficiency of indirect selection using the equation:

$$\frac{RS_p^s}{RS_p} = \frac{R_s^2 \times |r_{ps}|}{R_p^2}$$

where RS_p^s is the indirect response to selection; RS_p is the response to direct selection for the primary trait; R_s^2 is the plot level repeatability of the secondary trait; R_p^2 is the plot level repeatability of the primary trait; and r_{ps} is the genotypic correlation between the primary and secondary trait. Additionally, we estimated the coincidence of selection between primary and secondary traits based on rankings. To minimize the influence of the selection intensity, values ranging from 10 % to 50% ($IS = \{10\%, 15\%, \ldots, 50\%\}$) were utilized.

2.2.6. Phenomic prediction of primary traits (neural networks and partial least squares regression)

To determine the utility of vegetation indices, flight dates, and their specific combinations for predicting the primary traits, three prediction structures were analyzed regarding the utilization of the secondary traits:

1. A single index but with multiple flight dates as predictors;
2. Combinations of indices in a single-flight date as predictors; and
3. Spectral bands as predictors in a single-flight date.

These structures generated multiple prediction scenarios; For example, NDVI from flights at 69 and 128 DAS as predictors of AUDPC, or NDVI and MSR from the last flight as predictors of GY. The number of secondary traits utilized for predicting the primary trait varied depending on the use case. To accommodate these three structures, prediction methods included supervised machine-learning (ML) algorithms and partial least squares (PLSR).

The ML method used a neural network with resilient backpropagation and weight backtracking (RPROP algorithm; Riedmiller and Braun, 1994). The network structure was created using the *neuralnet* function (version 1.33) and was composed of an input layer with varying number of neurons depending on the scenario, two hidden layers of three and two neurons, and one output layer to be compared to the ground-truth (Figure 1; Günther and Fritsch, 2000). Predictors (input data) were used as the values of indices or spectral bands. The number of hidden layers and their neurons were selected based on preliminary tests (data not shown). A reduction of error between iterations of less than 1% (threshold of 0.01) as the stopping (convergence) criteria was administered. The entire training population was used in a single batch. Parameters of the model were then optimized using the *compute* function.

PLSR was also utilized for the prediction of primary traits when secondary traits were used as predictors. This method was implemented using the *pls* (version 2.7-0) R library (Mevik and Wehrens, 2015). The regression model was applied as follows:

$$\boldsymbol{y} = W\boldsymbol{\alpha} + \boldsymbol{\varepsilon}$$

where $\boldsymbol{y}$ is the vector of observations of the primary trait; α is the scalar/vector of regression coefficients of latent variables; and $\boldsymbol{\varepsilon}$: is the vector of errors [$\boldsymbol{\varepsilon} \sim N(0, I\sigma_\varepsilon^2)$]. W is the weight matrix relating the primary and secondary traits, which was obtained by the joint single value decomposition of primary and secondary traits. To determine the optimal number of latent variables (components), the permutation algorithm was utilized as described in Mevik and Wehrens (2015). These latent variables were utilized as predictors in the validation.

Prior to neural network and PLSR analyses, the data (primary and secondary traits) was normalized to 0-1 range using R's *scale* function. The assessment of relevant parameters to the predictive ability of the models in each scenario was completed using the following steps:

1. Random allocation of plots in training (75%; TS) and testing populations (25%, VS);
2. Train the models based on the training population information;
3. Prediction of testing population plots $\hat{\boldsymbol{y}} = \left\{\hat{y}_1, \ldots, \hat{y}_{n_{vs}}\right\}$;
4. Estimation of predictive ability ($r_{\hat{y}y}$; Pearson's product-moment correlation between predicted and observed values); and
5. Replication of steps 1 to 4 10 times $S_{i=1}^{b} = \{S_1, \ldots, S_{10}\}$ with common training and testing populations between scenarios, $S_b = \{VS_b, TS_b\}$, where b is the replication number.

The effect of methods and scenarios over predictive abilities was tested using analysis of variance (ANOVA), considering each repetition of the validation as a replication. To normalize the predictive abilities (correlations) prior to the analysis, Fisher's z transformation where by $z = \frac{1}{2} ln\left(\frac{1+r_{\hat{y}y}}{1-r_{\hat{y}y}}\right)$ was used. Additionally, Tukey's test was used to compare mean predictive abilities of scenarios using the *agricolae* version 1.2-8 library (De Mendiburu, 2016).

2.3. RESULTS

2.3.1. Exploratory analysis and data description

Anthracnose symptoms were minimal until 70 DAS after which there was a visual linear increase in incidence and severity until the last rating at 127 DAS (Figure 2). On a plot basis, anthracnose ratings ranged from zero (no presence of disease) to eight (severe disease) at 127 DAS. AUDPC values derived from these recordings varied from zero to 304.50 with an overall mean of 100.25 (Figure S1). GY ranged from 4.95 to 9.73, with a mean of 7.66 Mg ha^{-1}. Spatial dissemination of anthracnose progression can be observed from the recorded anthracnose incidence/severity scores over time (Figure S2). Hence, preliminary phenotypic analyses were carried out on primary and secondary traits to assess the influence of correlated errors to model fitting (Akaike criterion), which did not result in consistent modeling improvements.

Minimum, mean, and maximum values for spectral indices across all flights were respectively: CIg (2.11, 4.19 and 6.10); CIre (0.81, 2.22 and 4.15); MSR (0.56, 1.86 and 3.18); MSRre (0.48, 1.06 and 1.67); NDVI (0.32, 0.68 and 0.85); and NDVIre (0.29, 0.51 and 0.67). The amplitude of values increased during the crop cycle (Figure S1). However, most indices presented a decrease in mean value throughout the season. This same pattern was not present for the red edge-based indices, where a meaningful increase from 69 to 91 DAS was observed. Additionally, when this band was utilized for index calculation, lower mean values were obtained in each flight. The adjusted means of primary and secondary traits is also reported (Figure S3).

2.3.2. Factor significance testing and data quality

Based on the results from the phenotypic analysis, genotypic variation existed for all traits (Table S1). While replication partitioned some variation from the error term, the row and column factors were not significant. Flight date, a secondary trait-related source of variation, was significant for all evaluated indices.

Data quality information was assessed using the repeatabilities and experimental accuracy as metrics (Table 3). Repeatabilities on a plot basis were moderate for GY and high for anthracnose (AUDPC). Estimates of repeatability for secondary traits ranged from moderate to high, generally increasing in later stages of evaluation. Overall, the red edge-based indices had the highest values. Repeatabilities on an entry-mean basis were high for both primary and secondary traits. Experimental accuracy was high for primary and secondary traits regardless of flight date according to the scale proposed by Resende and Duarte (2007).

2.3.3. Indirect selection based on spectral indices (secondary traits)

Multi-trait (bivariate) modeling revealed genotypic correlations between primary and secondary traits (Figure 3). Spectral indices were positively correlated to GY and negatively correlated to AUDPC. Correlations of spectral indices to AUDPC increases in magnitude as the flight date neared crop maturity. For GY, values were highest at 102 and 119 DAS and decreased thereafter. Maximum correlations to AUDPC were observed at 119 DAS using NDVI (-0.89), and to GY at 102 DAS (0.69) using the same index. Genotypic correlations to AUDPC were generally higher than those to GY, especially after 91 DAS. None of the indices had higher correlations with any primary trait across all measured flight dates, suggesting specific index × flight date interactions.

Indirect selection using secondary traits appears to be effective in some situations. For GY, CIg (127 DAS flight), MSR, MSRre, NDVI and NDVIre (119 DAS) resulted in equal to/greater than 1.00 indirect selection to direct selection efficiency ratio (Table 4). For AUDPC, this was observed for CIre, MSR, MSRre, NDVI and NDVIre (119 DAS) and for MSR, MSRre, NDVI and NDVIre (127 DAS) (Table 4). Nevertheless, the genotypes selected by both methods were somewhat inconsistent (Figure S4). For example, no combination of index and flight resulted in clear superiority across all selection intensities.

2.3.4. Validation of prediction models

Predictive abilities obtained from the validation process are presented by index (Figure S5) and by flight (Figure S6) for primary traits. These values were regarded as a response variable and modeled in an analysis of variance with method, scenario (which is combinations of indices or flights), index or flight (index/flight), and double and triple interactions as fixed explanatory variables (Table 5). For clarity purposes, double and triple interactions are presented but not discussed in greater detail due to the number of levels.

For the "BY INDEX" validation scheme, the method, scenario, index, and method × index effects were statistically determinant in their predictive ability for AUDPC. For GY, similar factors were significant, in addition to the method × scenario interaction. Overall predictive ability means of indices varied from 0.62 to 0.75 for AUDPC and from 0.50 to 0.55 for GY (Table 6). Tukey's test revealed that NDVI and MSR were the best indices for

predicting AUDPC, while CIg was the worst predictor. For GY, all indices were statistically similar except for CIg, which had the lowest predictive ability. Statistical differences for GY were minimal for the tested scenarios, while more statistical differences were observed for AUDPC (Table S2). It is possible that information from early flights decreases predictive ability and should not be included in the model. Regarding the prediction methods, PLSR outperformed the machine learning method for predicting both AUDPC and GY (Table 7).

In the "BY FLIGHT" (Table 5) validation scheme, the method, scenario, flight, and method × flight effects were significant for the establishment of AUDPC predictive ability. For GY, flight and method × flight were the only significant factors. Predictive ability means of flights for GY ranged from 0.16 to 0.59, while for AUDPC, they varied from -0.07 to 0.81 (Table 8). Tukey mean comparisons tests indicated that 119 DAS and 102 DAS were the most useful in predicting AUDPC and GY, respectively. Scenarios (combination of spectral indices) examined resulted in statistically similar predictive abilities, except CIg alone for AUDPC, which had the lowest mean (Table S3). Concerning the prediction method, PLSR was more effective than the machine learning method for AUDPC but not for GY (Table 7).

2.4. DISCUSSION

2.4.1. Index development and assessment

Multiple scenarios were built to assess the optimum combinations of spectral indices and flight dates to predict GY and AUDPC. The predictive abilities of scenarios with either single or multiple indices were not different in their explanation and prediction of either GY and anthracnose resistance. This corroborates the findings of Li et al. (2018) on combinations of spectral indices for the prediction of sorghum biomass. Despite being developed for specific applications, the utilized spectral indices invariably target common traits such as chlorophyll content (Gitelson et al., 2003, 2005; Babar et al., 2006), LAI (Chen and Street, 1995; Xie et al., 2018), biomass (Rouse et al., 1973; Babar et al., 2006), and other health and yield component traits. These indices have different saturation points (Wu et al., 2008), with potentially different responses to optical/geometrical properties (Chen and Street, 1995), sensitivity to chlorophyll changes (Xie et al., 2018), and atmospheric/environmental influences (Myneni and Asrar, 1994). However, they did not increase the predictive ability when combined. Herein, indices were not complementary under the utilized methods/algorithms.

Hence, the most parsimonious option is to identify the single index that provides the higher prediction ability for each primary trait. For AUDPC, differences in predictive abilities of indices were larger than for GY, which is probably due to the higher repeatability of AUDPC. In validation, NDVI and MSR were superior for the prediction of AUDPC. This is consistent with reports by Pugh et al. (2018), who showed a strong linear relationship between NDVI and AUDPC scores. Huang et al. (2014) compared a set of indices for monitoring disease in wheat and found that NDVI and MSR were good at classifying healthy and diseased leaves (accuracies of 83.6% and 82.9%, respectively). Within the same species, NDVI has been reported in QTL mapping studies to be associated with spot blotch (Kumar et al., 2016) and stripe rust (Pretorius et al., 2017) resistance.

For GY, besides CIg, all indices presented similar predictive abilities. This index had the lowest values despite its higher genotypic correlation to the primary trait for two flight dates (119 and 127). The relationship between NDVI and GY in sorghum has previously been reported (Pugh et al., 2018). Moreover, Shafian et al. (2018)

found NDVI to be the best predictor over a range of indices. In maize, Torino et al. (2014) reported good results using NDVIre and CIre for predictions of GY. Thus, examples of the robustness of these indices at predicting GY with considerable ability are present. Although, it is the scientists' discretion to assess the usefulness and efficiency of this methodology when including it within their respective breeding program.

In this study, we employed multiple well-known established indices for the prediction of AUDPC and GY. Nevertheless, combined band values (by date) were also tested for the determination of their predictive ability of the evaluated primary traits (Figure S7). The highest means for AUDPC (0.48 at 119 in PLSR) and GY (0.57 at 127 in neural networks) were lower than the best scenarios presented when using indices (0.84 at 119 using NDVI/MSR in PLSR for AUDPC; 0.63 at 127 using MSRre in ML). Handcrafted indices resulted in better prediction abilities, even though the utilized methods are able to internally reproduce their equations/explain the variation. Studies conducted on other crops have reported that hyperspectral imagery was found to have better predictive abilities with bands than with individual indices, suggesting index estimation leads to information loss (Aguate et al., 2017; Montesinos-López et al., 2017). However, all bands available to our research were utilized herein for index estimation, which did not happen as reported in the aforementioned reports. Hence, additional studies in sorghum using hyperspectral data are warranted.

A series of secondary considerations can be drawn from this study that may be useful for researchers when working with vegetation indices in grain sorghum: *1)* plots should have at least two rows due to high interplot leaf overlap (Figure S8 A and B); *2)* whenever genotype is an explanatory variable in a primary-secondary trait correlation study (e.g. NDVI and GY), the number of treatments should be large enough so non-causal relationships are minimized; and *3)* both panicles and disease symptoms have lower index (e.g. NDVI) values than healthy leaves for all tested indices (Figure S8 C and D). Hence, the association of indices with GY/AUDPC - an association between plant health/vigor and yield/disease resistance - may be somewhat compromised. Therefore, one could use further image treatments and segmentation techniques (which was not in the scope of our study) to improve the correlation between primary traits and spectral indices.

2.4.2. On the best (combinations of) flight date(s)

As shown previously, prediction scenarios were built for determining the usefulness of remotely sensed temporal data for predicting GY and AUDPC. Results suggest that combining measurements of a spectral index from multiple flights does not result in a statistically significant increase in predictive ability for both primary traits. Aguate et al. (2017), working with hyperspectral data in maize, reported that the higher prediction abilities of GY were obtained when data from all the time points (5 flights) were combined. Zhou et al. (2017), estimating GY in rice, also found that combining flight dates results in increased correlation estimates. To our knowledge, however, no other result is available on the combination of dates using multispectral imagery for sorghum hybrids.

Regarding the dynamics of spectral indices, the amplitude of their adjusted means invariably increased from the first to the last evaluation date (Figure S3). This behavior matches the disease progress scores (Figure S2). Nevertheless, as genetic variances for spectral indices respectively increased throughout the growth cycle (Figure S9), the increment in genotypic correlations between primary and secondary traits (Figure 1) resulted from boosts in the genotypic covariance (Figure S10). Hence, as indirect selection relies on the repeatability of the primary and secondary trait and their covariance, better predictive abilities are expected at latter growth stages. The incremental

nature of the association between primary traits and spectral indices is a possible reason for the non-complementarity action of multiple flight dates.

In this sense, our results suggest that 102 DAS for GY and 119 DAS for AUDPC were the best dates to collect the secondary traits for prediction. These findings corroborate Pugh et al. (2018) for AUDPC yet contradict Shafian et al. (2018) for GY. The latter reported index assessment at the flowering stage to be the best estimator for this trait. However, GY reduction is a known disease symptom (Mofokeng et al., 2017). Hence, stronger associations between GY and secondary traits are expected when the widest range of disease resistance levels are present (final stages).

2.4.3. On the best prediction method

The PLSR is the method of choice for many remote sensing scientists due to its ability to deal with highly correlated parameters such as spectral bands and temporal data (Weber et al., 2012; Thorp et al., 2015; Aguate et al., 2017). Recently, interest in ML has increased in biological fields due to its ability to extract hidden patterns/structures, learn from data sets, and adapt to iteratively increase predictive ability (Angermueller et al., 2016; Liakos et al., 2018). In this study, the PLSR, in most cases, was more effective than ML in predicting both GY and anthracnose incidence. Additionally, best results were achieved with single flight dates and indices. It is worth noting that, in these cases, PSLR is equivalent to a conventional linear model. Superiority of the benchmark model was also reported by Montesinos-López et al. (2018) where the ML method was inferior to GBLUP. Given that the accuracy of neural networks potentially improves as the number of observations increase, additional studies with different parameters and higher number of observations should be completed to further elucidate the effective use of ML methods.

2.4.4. Application of remote sensing data for plant breeding

Results presented herein indicate that both GY and foliar disease ratings (mostly due to anthracnose) can be estimated using HTP traits with considerable accuracy. In addition, we also demonstrate that the ability to select depends on spectral indices and specific flight dates. However, the difference in mean predictive ability of GY from the flights at 102 (best for GY) and 119 DAS (best for AUDPC) was only 0.02. Thus, the prediction of both GY and AUDPC using a single flight at 119 DAS and a single index (NDVI or MSR) should suffice and would potentially increase selection efficiency.

Increasing gain from selection over time is an important goal of any breeding program. Indirect selection should be considered when it is faster or cheaper than direct selection. In this study, the most predictive flight date (119 DAS) combined with the best general indices (NDVI and MSR), yielded selection efficiencies of ~1.03 for GY and ~1.02 for AUDPC. In this situation, selection based on secondary traits may lead to greater genetic gain compared to direct selection on the primary trait. Additionally, further increases in the efficiency of a breeding program may be achieved when one considers the relative cost of HTP for phenotyping, the advantage of earlier selection windows, reduction of subjectivity in measurements, and its ability to exploit genetic variability by increasing the effective size of the breeding population (Araus et al., 2018).

Further breeding inferences rely on the coincidence of selection. Herein there was a moderate correlation between AUDPC and GY (-0.48) and only a few hybrids were best for AUDPC and GY at high selection intensities (Figure S11). Thus, while high yield was associated with disease resistance, not all genotypes with good disease resistance had high GY. Nonetheless, the coincidence of selection between spectral indices increases as selection intensity decreases, following similar patterns previously presented on regarding genomic prediction studies (Matias et al., 2017; Galli et al., 2018). In this sense, spectral indices may serve as a form of phenomic selection where, after optimization, selection of hybrids could be based on a series of remotely sensed traits associated with breeding values in the same way that genomic selection is applied to genotypes. Like genomic selection, phenomic selection is likely more useful at eliminating undesirable genotypes than for selecting superior genotypes. At last, this study showed the possibility of optimizing data collection and processing in order to increase the efficiency of new phenotyping methodologies in grain sorghum, suggesting that it can also be performed in other crops.

REFERENCES

Aguate, F.M., S. Trachsel, L.G. Pérez, J. Burgueño, J. Crossa, et al. 2017. Use of Hyperspectral Image Data Outperforms Vegetation Indices in Prediction of Maize Yield. Crop Sci. 0(0): 0. doi: 10.2135/cropsci2017.01.0007.

Angermueller, C., T. Pärnamaa, L. Parts, and O. Stegle. 2016. Deep learning for computational biology. Mol. Syst. Biol. 12(7): 878. doi: 10.15252/msb.20156651.

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

Araus, J.L., and S.C. Kefauver. 2018. Breeding to adapt agriculture to climate change: affordable phenotyping solutions. Curr. Opin. Plant Biol. doi: 10.1016/j.pbi.2018.05.003.

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.

Babar, M.A., M.P. Reynolds, M. Van Ginkel, A.R. Klatt, W.R. Raun, et al. 2006. Spectral reflectance to estimate genetic variation for in-season biomass, leaf chlorophyll, and canopy temperature in wheat. Crop Sci. 46(3): 1046–1057. doi: 10.2135/cropsci2005.0211.

Boyles, R.E., B.K. Pfeiffer, E.A. Cooper, K.J. Zielinski, M.T. Myers, et al. 2017. Quantitative trait loci mapping of agronomic and yield traits in two grain sorghum biparental families. Crop Sci. doi: 10.2135/cropsci2016.12.0988.

Butler, D., B. Cullis, A. Gilmour, and B. Gogel. 2007. Analysis of Mixed Models for S–language Environments: ASReml–R Reference Manual. (March).

Chen, J.M. 1996. Evaluation of vegetation indices and a modified simple ratio for boreal applications. Can. J. Remote Sens. doi: 10.1080/07038992.1996.10855178.

Chen, J.M., and B. Street. 1995. Evaluation of Vegetation Indices and a Modified Simple Ratio for Boreal Applications. Can. J. Remote Sens. (613): 1–21.

Cobb, J.N., G. DeClerck, A. Greenberg, R. Clark, and S. McCouch. 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(4): 867–887. doi: 10.1007/s00122-013-2066-0.

da Costa, R. V., L. Zambolim, L. V. Cota, D.D. da Silva, J.A.S. Rodrigues, et al. 2011. Genetic control of sorghum resistance to leaf anthracnose. Plant Pathol. 60(6): 1162–1168. doi: 10.1111/j.1365-3059.2011.02469.x.

Feng, H., Z. Guo, W. Yang, C. Huang, G. Chen, et al. 2017. An integrated hyperspectral imaging and genome-wide association analysis platform provides spectral and genetic insights into the natural variation in rice. Sci. Rep. 7(1): 4401. doi: 10.1038/s41598-017-04668-8.

Galli, G., D.H. Lyra, F.C. Alves, Í.S.C. Granato, M.B. e Sousa, et al. 2018. Impact of Phenotypic Correction Method and Missing Phenotypic Data on Genomic Prediction of Maize Hybrids. Crop Sci. 58(4): 1481. doi: 10.2135/cropsci2017.07.0459.

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

Gitelson, A., and M.N. Merzlyak. 1994. Spectral Reflectance Changes Associated with Autumn Senescence of Aesculus hippocastanum L. and Acer platanoides L. Leaves. Spectral Features and Relation to Chlorophyll Estimation. J. Plant Physiol. doi: 10.1016/S0176-1617(11)81633-0.

Gitelson, A.A., A. Viña, V. Ciganda, D.C. Rundquist, and T.J. Arkebauer. 2005. Remote estimation of canopy chlorophyll content in crops. Geophys. Res. Lett. doi: 10.1029/2005GL022688.

GRASS Development Team. 2018. Geographic Resources Analysis Support System (GRASS) Software, Version 7.2.

Günther, F., and S. Fritsch. 2000. Training of Neural Networks. p. 51–63

Henderson, C.R. 1975. Best linear unbiased estimation and prediction under a selection model published by : international biometric society stable. Biometrics 31(2): 423–447.

Hijmans, R.J., J. v. Etten, J. Cheng, M. Mattiuzzi, M. Sumner, et al. 2015. Package ' raster .' https://cran.r-project.org/web/packages/raster/raster.pdf.

Huang, W., Q. Guan, J. Luo, J. Zhang, J. Zhao, et al. 2014. New Optimized Spectral Indices for Identifying and Monitoring Winter Wheat Diseases. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 7(6): 2516–2524. doi: 10.1109/JSTARS.2013.2294961.

Kumar, S., M.S. Röder, R.P. Singh, S. Kumar, R. Chand, et al. 2016. Mapping of spot blotch disease resistance using NDVI as a substitute to visual observation in wheat (Triticum aestivum L.). Mol. Breed. 36(7). doi: 10.1007/s11032-016-0515-6.

Li, J., and Y. Shi. 2018. Elucidating Sorghum Biomass , Nitrogen and Chlorophyll Contents With Spectral and Morphological Traits Derived From Unmanned Aircraft System. 9(October): 1–12. doi: 10.3389/fpls.2018.01406.

Liakos, K.G., P. Busato, D. Moshou, S. Pearson, and D. Bochtis. 2018. Machine learning in agriculture: A review. Sensors (Switzerland) 18(8): 1–29. doi: 10.3390/s18082674.

Matias, F.I., G. Galli, I.S. Correia Granato, and R. Fritsche-Neto. 2017. Genomic Prediction of Autogamous and Allogamous Plants by SNPs and Haplotypes. Crop Sci. 57(6): 2951. doi: 10.2135/cropsci2017.01.0022.

De Mendiburu, F. 2016. Package 'agricolae.' Stat. Proced. Agric. Res.

Mevik, B.-H., and R. Wehrens. 2015. The pls Package: Principal Component and Partial Least Squares Regression in R . J. Stat. Softw. doi: 10.18637/jss.v018.i02.

Mofokeng, M.A., H. Shimelis, M. Laing, and N. Shargie. 2017. Sorghum [Sorghum bicolor (L.) Moench] breeding for resistance to leaf and stalk anthracnose, Colletotrichum sublineolum, and improved yield: Progress and prospects. Aust. J. Crop Sci. 11(9): 1078–1085. doi: 10.21475/ajcs.17.11.09.pne347.

Montesinos-López, O.A., A. Montesinos-López, J. Crossa, G. los Campos, G. Alvarado, et al. 2017. Predicting grain yield using canopy hyperspectral reflectance in wheat breeding data. Plant Methods 13(1): 1–23. doi: 10.1186/s13007-016-0154-2.

Montesinos-López, A., O.A. Montesinos-López, D. Gianola, J. Crossa, and C.M. Hernández-Suárez. 2018. Multi-environment Genomic Prediction of Plant Traits Using Deep Learners with Dense Architecture. G3: Genes|Genomes|Genetics 8(December): g3.200740.2018. doi: 10.1534/g3.118.200740.

Myneni, R.B., and G. Asrar. 1994. Atmospheric effects and spectral vegetation indices. Remote Sens. Environ. 47(3): 390–402. doi: 10.1016/0034-4257(94)90106-6.

Patil, N.Y., R.R. Klein, C.L. Williams, S.D. Collins, J.E. Knoll, et al. 2017. Quantitative trait loci associated with anthracnose resistance in sorghum. Crop Sci. 57(2): 877–890. doi: 10.2135/cropsci2016.09.0793.

Pebesma, E.J., and R.S. Bivand. 2005. Classes and methods for spatial data in R. R news.

Potgieter, A.B., B. George-Jaeggli, S.C. Chapman, K. Laws, L.A. Suárez Cadavid, et al. 2017. Multi-Spectral Imaging from an Unmanned Aerial Vehicle Enables the Assessment of Seasonal Leaf Area Dynamics of Sorghum Breeding Lines. Front. Plant Sci. 8(September): 1–11. doi: 10.3389/fpls.2017.01532.

Pretorius, Z.A., C.X. Lan, R. Prins, V. Knight, N.W. McLaren, et al. 2017. Application of remote sensing to identify adult plant resistance loci to stripe rust in two bread wheat mapping populations. Precis. Agric. 18(4): 411–428. doi: 10.1007/s11119-016-9461-x.

Pugh, N.A., X. Han, S.D. Collins, J.A. Thomasson, D. Cope, et al. 2018. Estimation of plant health in a sorghum field infected with anthracnose using a fixed-wing unmanned aerial system. J. Crop Improv. 32(6): 861–877. doi: 10.1080/15427528.2018.1535462.

QGIS Development Team. 2018. QGIS Geographic Information System. Open Source Geospatial Found. Proj. doi: 10.1038/nrg1769.

R Core Team. 2011. R Development Core Team. R A Lang. Environ. Stat. Comput.

Resende, M.D.V. De, and J.B. Duarte. 2007. Precisão e controle de qualidade em experimentos de avaliação de cultivares. Pesqui. Agropecuária Trop. 37(3): 182–194.

Riedmiller, M., and H. Braun. 1994. A direct adaptive method for faster backpropagation learning: the RPROP algorithm. IEEE International Conference on Neural Networks. IEEE. p. 586–591

Rouse, J.W., R.H. Hass, J.A. Schell, and D.W. Deering. 1973. Monitoring vegetation systems in the great plains with ERTS. Third Earth Resour. Technol. Satell. Symp. 1: 309–317. doi: citeulike-article-id:12009708.

Shafian, S., N. Rajan, R. Schnell, M. Bagavathiannan, J. Valasek, et al. 2018. Unmanned aerial systems-based remote sensing for monitoring sorghum growth and development. PLoS One 13(5). doi: 10.1371/journal.pone.0196605.

Shi, S., F.I. Azam, H. Li, X. Chang, B. Li, et al. 2017. Mapping QTL for stay-green and agronomic traits in wheat under diverse water regimes. Euphytica 213(11): 1–19. doi: 10.1007/s10681-017-2002-5.

Stanton, C., M.J. Starek, N. Elliott, M. Brewer, M.M. Maeda, et al. 2017. Unmanned aircraft system-derived crop height and normalized difference vegetation index metrics for sorghum yield and aphid stress assessment. J. Appl. Remote Sens. 11(2): 026035. doi: 10.1117/1.JRS.11.026035.

Thorp, K.R., M.A. Gore, P. Andrade-Sanchez, A.E. Carmo-Silva, S.M. Welch, et al. 2015. Proximal hyperspectral sensing and data analysis approaches for field-based plant phenomics. Comput. Electron. Agric. 118: 225–236. doi: 10.1016/j.compag.2015.09.005.

Torino, M.S., B. V. Ortiz, J.P. Fulton, K.S. Balkcom, and C.W. Wood. 2014. Evaluation of vegetation indices for early assessment of corn status and yield potential in the southeastern United States. Agron. J. 106(4): 1389–1401. doi: 10.2134/agronj13.0578.

Weber, V.S., J.L. Araus, J.E. Cairns, C. Sanchez, A.E. Melchinger, et al. 2012. Prediction of grain yield using reflectance spectra of canopy and leaves in maize plants grown under different water regimes. F. Crop. Res. 128: 82–90. doi: 10.1016/j.fcr.2011.12.016.

Wu, C., Z. Niu, Q. Tang, and W. Huang. 2008. Estimating chlorophyll content from hyperspectral vegetation indices: Modeling and validation. Agric. For. Meteorol. 148(8–9): 1230–1241. doi: 10.1016/j.agrformet.2008.03.005.

Xie, Q., J. Dash, W. Huang, D. Peng, Q. Qin, et al. 2018. Vegetation Indices Combining the Red and Red-Edge Spectral Information for Leaf Area Index Retrieval. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 11(5): 1482–1492. doi: 10.1109/JSTARS.2018.2813281.

Zhou, X., H.B. Zheng, X.Q. Xu, J.Y. He, X.K. Ge, et al. 2017. Predicting grain yield in rice using multi-temporal vegetation indices from UAV-based multispectral and digital imagery. ISPRS J. Photogramm. Remote Sens. 130: 246–255. doi: 10.1016/j.isprsjprs.2017.05.003.

TABLES

Table 1. Days after sowing (DAS) at which flight was performed, sensor, altitude (relative to ground), ground resolution of orthomosaic, number of GCPs, and accuracy of five flights over the sorghum trial in College Station, TX. Ground resolution and accuracy were estimated in Agisoft Photoscan Pro averaged over all aligned images.

DAS	Sensor	Altitude (m)	Ground resolution (cm px^{-1})	Number of GCPs	Accuracy (cm)
69	MicaSense® RedEdge	30.30	2.00	8	0.46
91	SlantRange® 3P	49.60	1.96	8	0.73
102	SlantRange® 3P	50.90	2.01	8	1.75
119	SlantRange® 3P	50.00	1.98	8	0.60
127	SlantRange® 3P	50.80	2.01	7	0.89

Table 2. Utilized vegetation indices based on green (ρ_G), red (ρ_R), red edge (ρ_{RE}), and near infra-red (ρ_{NIR}) bands.

Index	Acronym	Formula
Chlorophyll Index[1]	CIg	$\frac{\rho_{NIR}}{\rho_G} - 1$
Red edge chlorophyll index[2]	CIre	$\frac{\rho_{NIR}}{\rho_{RE}} - 1$
Modified simple ratio[3]	MSR	$\frac{\rho_{NIR}/\rho_R - 1}{\sqrt{\rho_{NIR}/\rho_R + 1}}$
Red edge modified simple ratio[4]	MSRre	$\frac{\rho_{NIR}/\rho_{RE} - 1}{\sqrt{\rho_{NIR}/\rho_{RE} + 1}}$
Normalized difference vegetation index[5]	NDVI	$\frac{\rho_{NIR} - \rho_R}{\rho_{NIR} + \rho_R}$
Red edge normalized difference vegetation index[6]	NDVIre	$\frac{\rho_{NIR} - \rho_{RE}}{\rho_{NIR} + \rho_{RE}}$

[1]Gitelson et al. (2003); [2]Gitelson et al. (2005); [3]Chen (1996); [4]Wu et al. (2008); [5]Rouse et al. (1973); [6]Gitelson and Merzlyak (1994).

Table 3. Repeatabilities at plot and entry-mean level and experimental accuracy of anthracnose area under disease progress curve (AUDPC), grain yield (GY), chlorophyll index (CIg), modified simple ratio (MSR), normalized difference vegetation index (NDVI), and their red edge counterparts (CIre, MSRre and NDVIre, respectively) in sorghum hybrids. Repeatability values for secondary traits are presented across five different flights shown in days after sowing (DAS).

	Plot level (DAS)					**Entry-mean level* (DAS)**					**Accuracy* (DAS)**				
Trait	**69**	**91**	**102**	**119**	**127**	**69**	**91**	**102**	**119**	**127**	**69**	**91**	**102**	**119**	**127**
CIg	0.49	0.69	0.85	0.85	0.89	0.75	0.87	0.94	0.94	0.96	0.86	0.93	0.97	0.97	0.98
CIre	0.49	0.77	0.81	0.93	0.91	0.74	0.91	0.93	0.98	0.97	0.86	0.95	0.96	0.99	0.98
MSR	0.47	0.62	0.82	0.91	0.90	0.72	0.83	0.93	0.97	0.96	0.85	0.91	0.97	0.98	0.98
MSRre	0.49	0.77	0.82	0.93	0.91	0.74	0.91	0.93	0.98	0.97	0.86	0.95	0.97	0.99	0.98
NDVI	0.46	0.63	0.81	0.89	0.90	0.72	0.83	0.93	0.96	0.96	0.85	0.91	0.96	0.98	0.98
NDVIre	0.48	0.76	0.82	0.93	0.91	0.74	0.91	0.93	0.97	0.97	0.86	0.95	0.97	0.99	0.98
GY	-	-	-	-	0.63	-	-	-	-	0.84	-	-	-	-	0.91
AUDPC	-	-	-	-	0.80	-	-	-	-	0.92	-	-	-	-	0.96

*Based on 3 replications.

Table 4. Efficiency of indirect selection for hybrid sorghum grain yield (GY) and anthracnose area under disease progress curve (AUDPC) using vegetation indices (CIg, CIre, MSR, MSRre, NDVI, and NDVIre) calculated from five flights (69, 91, 102, 119 and 127 days after sowing) in College Station, TX.

	GY					**AUDPC**				
Index	**69**	**91**	**102**	**119**	**127**	**69**	**91**	**102**	**119**	**127**
CIg	0.22	0.70	0.96	0.99	1.02	0.16	0.56	0.72	0.78	0.88
CIre	0.40	0.80	0.93	0.99	0.90	0.17	0.65	0.84	1.01	0.99
MSR	0.20	0.64	0.98	1.03	0.98	0.15	0.59	0.85	1.02	1.01
MSRre	0.40	0.80	0.94	1.01	0.91	0.17	0.65	0.84	1.02	1.00
NDVI	0.19	0.64	0.97	1.03	0.99	0.14	0.61	0.84	1.01	1.02
NDVIre	0.39	0.79	0.95	1.03	0.93	0.18	0.65	0.84	1.03	1.00

Table 5. Three-way analysis of variance on predictive abilities using method (PLSR and ML), scenario (indices and their combinations or flights and their combinations), index/flight date, and their double and triple interactions.

	BY INDEX					BY FLIGHT				
Source	**Df**	**AUDPC**[F]		**GY**		**Df**	**AUDPC**		**GY**	
Method	1	4.21	***	6.88	***	1	0.93	***	0.04	NS
Scenario	30	6.73	***	1.02	***	62	0.10	***	0.01	NS
Index/Flight	5	7.36	***	0.39	***	4	323.24	***	61.98	***
Methods × Scenario	30	0.04	NS	0.13	***	62	0.01	NS	0.01	NS
Method × Index/Flight	5	0.10	**	0.07	*	4	0.46	***	0.16	***
Scenario × Index/Flight	150	0.03	NS	0.02	NS	248	0.02	NS	0.01	NS
Method × Scenario × Index/Flight	150	0.01	NS	0.01	NS	248	0.01	NS	0.01	NS
Residuals	3348	0.03		0.03		5670	0.03		0.03	

Df: degrees of freedom;
[F]Mean square value and significance codes in probability level ranges: *** < 0.001 ≤ ** < 0.01 ≤ * < 0.05 ≤ NS.

Table 6. Tukey's HSD test based on ten-replication validation predictive abilities for anthracnose area under disease progress curve (AUDPC) and grain yield (GY) using six spectral indices (CIg, CIre, MSR, MSRre, NDVI, and NDVIre) on sorghum hybrids in College Station, TX. Values herein are presented as means with their appropriate statistical grouping.

Index	**AUDPC**		**GY**	
NDVI	0.75	a*	0.55	a
MSR	0.75	a	0.55	a
NDVIre	0.73	b	0.55	a
MSRre	0.72	b	0.54	a
CIre	0.72	b	0.54	a
CIg	0.62	c	0.50	b

*groups with different letters are statistically different at 0.05 probability.

Table 7. Tukey's HSD test based on ten-replication validation predictive abilities of anthracnose area under disease progress curve (AUDPC) and grain yield (GY) using two prediction methods (Machine Learning - ML and Partial Least Squares Regression - PLSR) on sorghum hybrids in College Station, TX. Values herein are presented as means with their appropriate statistical grouping.

	BY INDEX		BY FLIGHT	
Method	**AUDPC**	**GY**	**AUDPC**	**GY**
PLSR	0.73 a	0.57 a*	0.57 a	0.49 a
ML	0.70 b	0.51 b	0.56 b	0.48 a

*groups with different letters are statistically different at 0.05 probability.

Table 8. Tukey's HSD test based on ten-replication validation predictive abilities for anthracnose area under disease progress curve (AUDPC) and grain yield (GY) using five flights (69, 91, 102, 119, and 127 days after sowing) on sorghum hybrids in College Station, TX. Values are shown as means and letters are their respective statistical groupings.

Flight (DAS)	**AUDPC**	**GY**
127	0.79 b	0.58 b
119	0.81 a	0.57 b
102	0.73 c	0.59 a
91	0.54 d	0.52 c
69	-0.07 e	0.16 d

*groups with different letters are statistically different at 0.05 probability.

FIGURES

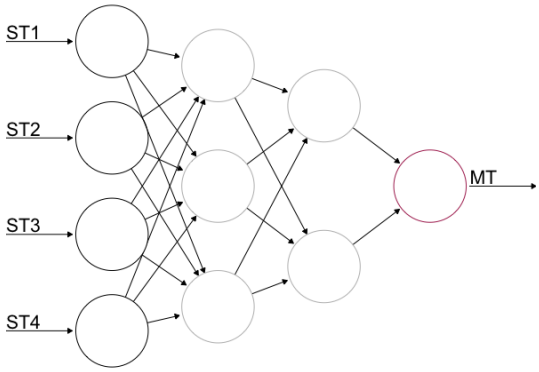

Figure 1. Example structure of the utilized neural network. The structure is composed of an input layer (predictors; black circles), two hidden layers of three and two neurons (gray), and one output layer (red). Input data (ST1, ST2, ST3, ST4) are the values of indices or spectral bands of a plot. Output data (MT) is the network prediction of a primary trait.

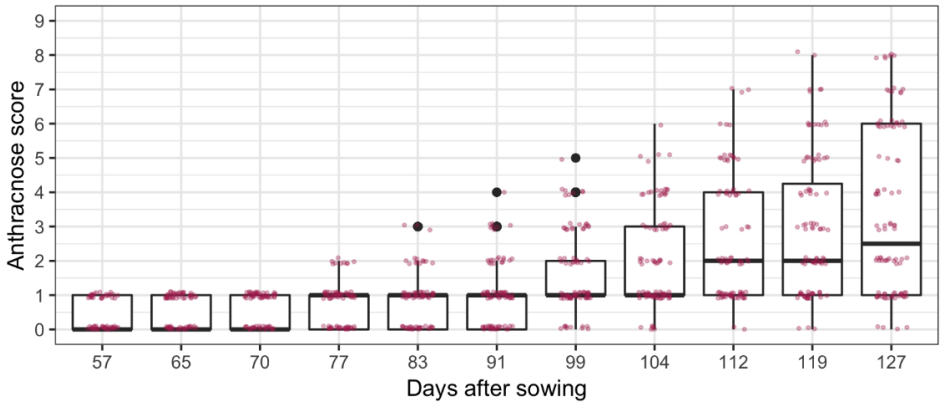

Figure 2. Boxplot of anthracnose scores on sorghum hybrids on 11 time points in College Station, TX. Maroon points are the values of plots and solid black points are the scores considered outliers.

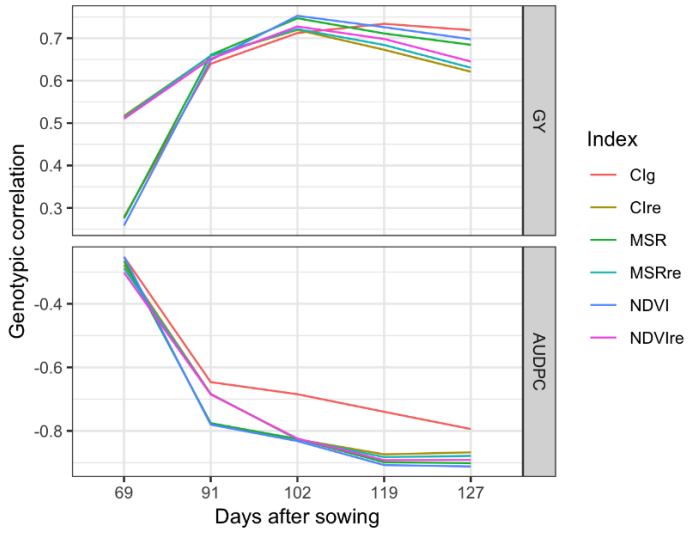

Figure 3. Genotypic correlation from bivariate mixed model of primary (grain yield - GY; and anthracnose area under disease progress curve - AUDPC) and secondary traits (chlorophyll index - CIg; modified simple ratio - MSR; normalized difference vegetation index - NDVI; red edge chlorophyll index - CIre; red edge modified simple ratio - MSRre; and normalized difference vegetation index - NDVIre) of sorghum hybrids across five flight dates (69, 91, 102, 119 and 127 days after sowing) in College Station, TX.

SUPPLEMENTARY TABLES

Table S1. Likelihood-ratio test (LRT) of random (genotype, row within replication, and column) and Wald statistics of fixed effects (flight date and replication) for anthracnose area under disease progress curve (AUDPC), grain yield (GY), chlorophyll index (CIg), modified simple ratio (MSR), normalized difference vegetation index (NDVI), and their red edge counterparts (CIre, MSRre and NDVIre, respectively) in sorghum hybrids.

Source	AUDPC	GY	CIg	CIre	MSR	MSRre	NDVI	NDVIre
Genotype	67.96 ***	33.63 ***	197.88 ***	216.81 ***	179.82 ***	213.54 ***	162.78 ***	195.70 ***
Row [Rep]	0.00 NS	0.27 NS	2.21 NS	1.81 NS	0.68 NS	0.82 NS	0.00 NS	0.05 NS
Column	0.20 NS	0.16 NS	0.00 NS	0.04 NS	0.00 NS	0.00 NS	0.00 NS	0.00 NS
Flight date	-	-	1092.27 ***	2831.77 ***	5468.99 ***	2487.02 ***	2276.31 ***	1905.37 ***
Replication	14.18 ***	10.51 **	4.37 NS	16.48 ***	8.07 *	17.11 ***	8.58 *	17.99 ***

Significance codes in probability level ranges: *** < 0.001 ≤ ** < 0.01 ≤ * < 0.05 ≤ NS.

Table S2. Tukey's honestly significant difference test based on ten-replication validation predictive abilities of grain yield (GY) and area under disease progress curve (AUDPC) using combination of flights (69, 91, 102, 119, and 127 days after sowing) on sorghum hybrids in College Station, TX. Presented values are means and letters are groups.

Dates	AUDPC		Dates	GY	
119	0.80	a*	102	0.580	a
119;127	0.79	ab	119	0.580	a
69;119	0.79	ab	127	0.580	a
91;119;127	0.78	abc	91;102	0.570	a
127	0.78	abc	119;127	0.570	a
102;119;127	0.78	abc	102;127	0.570	a
69;119;127	0.78	abc	102;119	0.570	a
102;119	0.78	abc	91;102;119;127	0.560	a
91;102;119;127	0.78	abc	91;102;127	0.560	a
91;119	0.78	abc	69;127	0.560	a
69;102;119;127	0.78	abcd	91;127	0.560	a
102;127	0.78	abcd	91;119	0.560	a
69;91;119;127	0.77	abcd	91;119;127	0.560	ab
69;102;119	0.77	abcd	69;102;127	0.560	ab
69;127	0.77	abcd	69;102	0.560	ab
69;91;102;119;127	0.77	abcd	69;102;119;127	0.560	ab
69;91;102;119	0.77	abcd	102;119;127	0.560	ab
69;102;127	0.76	abcd	69;119	0.550	ab
91;127	0.76	abcd	69;91;102;127	0.550	ab
91;102;119	0.76	abcd	69;91;127	0.550	ab
69;91;119	0.76	bcd	91;102;119	0.540	ab
91;102;127	0.75	cd	69;91;119	0.550	ab
69;91;102;127	0.75	cd	69;119;127	0.550	ab
69;91;127	0.74	de	69;91;102	0.540	ab
102	0.71	ef	69;91;119;127	0.540	ab
69;102	0.70	fg	69;102;119	0.540	ab
91;102	0.68	fg	69;91;102;119	0.530	ab
69;91;102	0.66	g	69;91;102;119;127	0.530	ab
91	0.53	h	91	0.500	bc
69;91	0.47	h	69;91	0.440	c
69	-0.07	i	69	0.160	d

*groups with different letters are statistically different at 0.05 probability.

Table S3. Summary of Tukey's honestly significant difference test based on ten-replication validation predictive abilities of grain yield (GY) and area under disease progress curve (AUDPC) using six spectral indices (CIg, CIre, MSR, MSRre, NDVI, and NDVIre) on sorghum hybrids in College Station, TX. Presented values are means and letters are groups.

Indices	AUDPC		Indices	GY	
MSR;NDVI	0.58	a*	NDVI;NDVIre	0.500	a
NDVI;NDVIre	0.58	a	NDVI	0.500	a
CIre;MSR;NDVI	0.58	a	MSR;NDVI;NDVIre	0.500	a
CIre;NDVI	0.58	a	MSR;NDVIre	0.500	a
MSR;NDVIre	0.58	a	MSR;NDVI	0.500	a
MSR;MSRre	0.58	a	MSR	0.500	a
MSR;MSRre;NDVI;NDVIre	0.58	a	MSR;MSRre;NDVI;NDVIre	0.490	a
MSR;MSRre;NDVI	0.58	a	MSRre;NDVI	0.490	a
CIre;MSR	0.58	a	CIre;MSR;NDVI;NDVIre	0.490	a
CIre;NDVI;NDVIre	0.58	a	CIre;MSR	0.490	a
CIre;MSRre;NDVI	0.58	a	MSR;MSRre;NDVI	0.490	a
MSR;MSRre;NDVIre	0.58	a	CIre;MSR;NDVI	0.490	a
CIre;MSR;MSRre;NDVI;NDVIre	0.58	a	CIre;MSR;MSRre;NDVIre	0.490	a
MSRre;NDVI;NDVIre	0.58	a	CIre;MSR;MSRre;NDVI	0.490	a
NDVI	0.57	a	CIre;MSRre;NDVI;NDVIre	0.490	a
MSR	0.57	a	MSR;MSRre	0.490	a
			…		
CIg;CIre;MSRre;NDVI;NDVIre	0.55	a	CIg;NDVIre	0.480	a
CIg;MSR	0.55	a	CIg;MSRre	0.480	a
CIg;CIre;NDVI	0.55	a	CIre;MSRre	0.480	a
CIg;CIre;MSR;MSRre;NDVI;NDVIre	0.55	a	CIg;CIre;MSRre;NDVIre	0.480	a
CIg;MSR;MSRre	0.55	a	CIg;CIre;MSR;NDVI;NDVIre	0.480	a
CIg;CIre;MSR;MSRre;NDVIre	0.55	a	CIg;MSR;NDVI;NDVIre	0.480	a
CIg;MSRre;NDVIre	0.55	a	CIre	0.480	a
CIg;CIre;NDVIre	0.55	a	CIg;CIre;MSR;MSRre;NDVIre	0.480	a
CIg;MSR;NDVIre	0.54	a	CIg;CIre	0.480	a
CIg;NDVI;NDVIre	0.54	a	CIg;CIre;MSR;MSRre	0.480	a
CIg;CIre;MSRre	0.54	a	CIg;CIre;MSR;NDVIre	0.470	a
CIg;CIre;MSRre;NDVIre	0.54	a	CIg;NDVI	0.470	a
CIg;NDVIre	0.54	a	CIg;MSR;MSRre	0.480	a
CIg;CIre	0.54	a	CIg;MSR	0.470	a
CIg;MSRre	0.54	a	CIg;CIre;MSR;MSRre;NDVI	0.470	a
CIg	0.47	b	CIg	0.460	a

*groups with different letters are statistically different at 0.05 probability.

SUPPLEMENTARY FIGURES

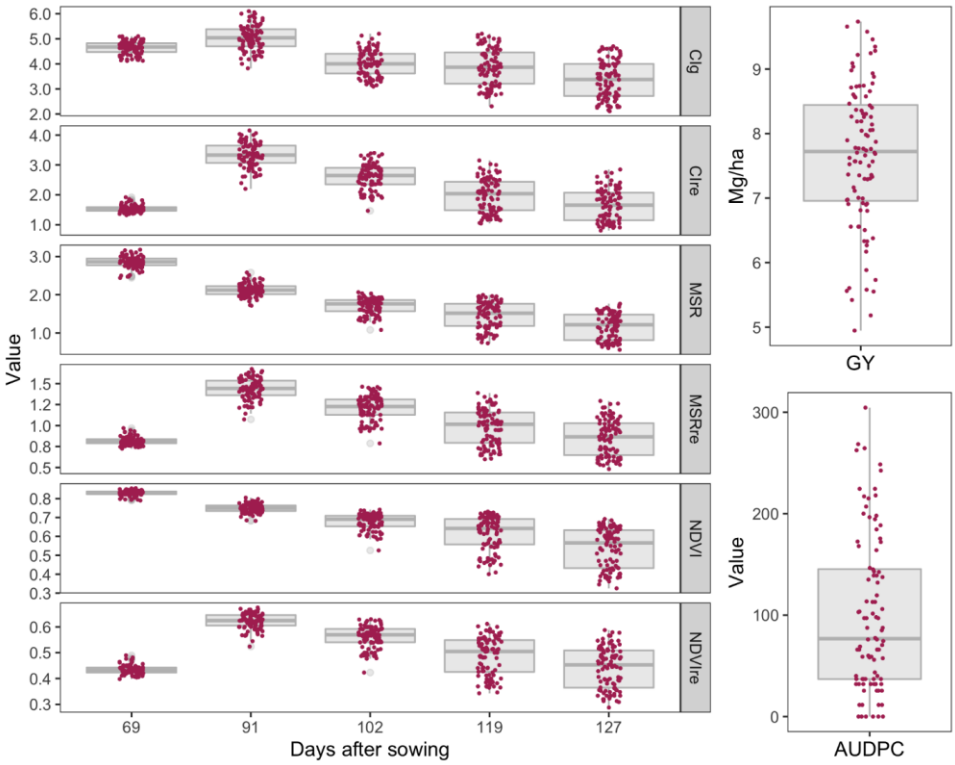

Figure S1. Boxplots of plot values of grain yield (GY), area under disease progress curve (AUDPC), chlorophyll index (CIg), modified simple ratio (MSR), normalized difference vegetation index (NDVI), and their formulas based on the red edge (CIre, MSRre and NDVIre, respectively) in sorghum hybrids on five flight dates in College Station, TX.

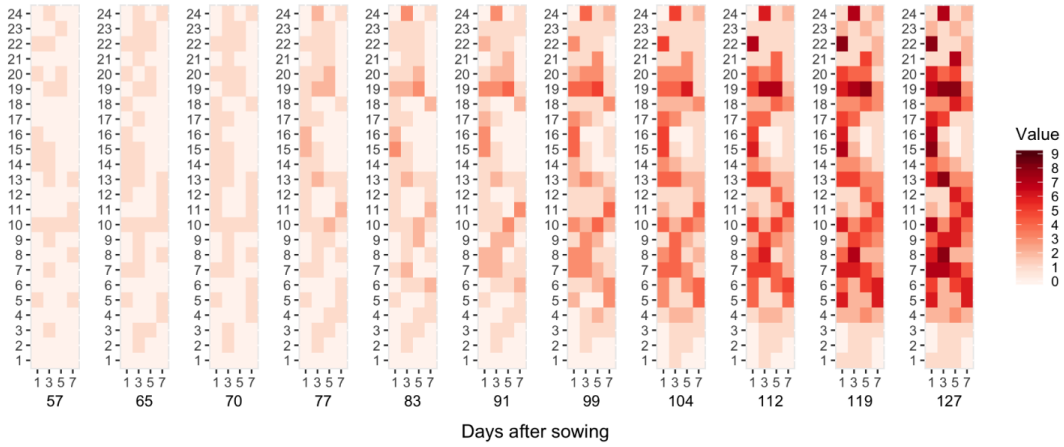

Figure S2. Experimental area of sorghum hybrids inoculated with anthracnose (*C. sublineolum*) in College Station, TX. Experiment was laid out in 24 rows and four columns (two column plots). Graphs represent the progress of visual scores across 11 time points from 57 to 127 days after sowing.

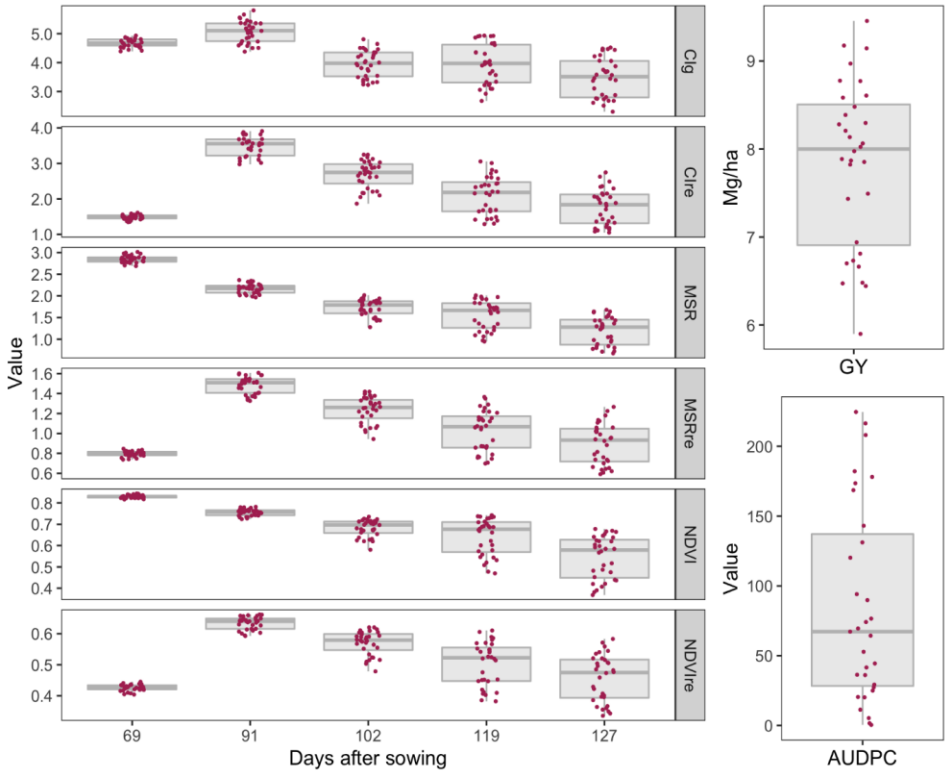

Figure S3. Boxplots of genotype adjusted means of grain yield (GY), area under disease progress curve (AUDPC), chlorophyll index (CIg), modified simple ratio (MSR), normalized difference vegetation index (NDVI), and their formulas based on the red edge (CIre, MSRre and NDVIre, respectively) in sorghum hybrids on five flight dates in College Station, TX.

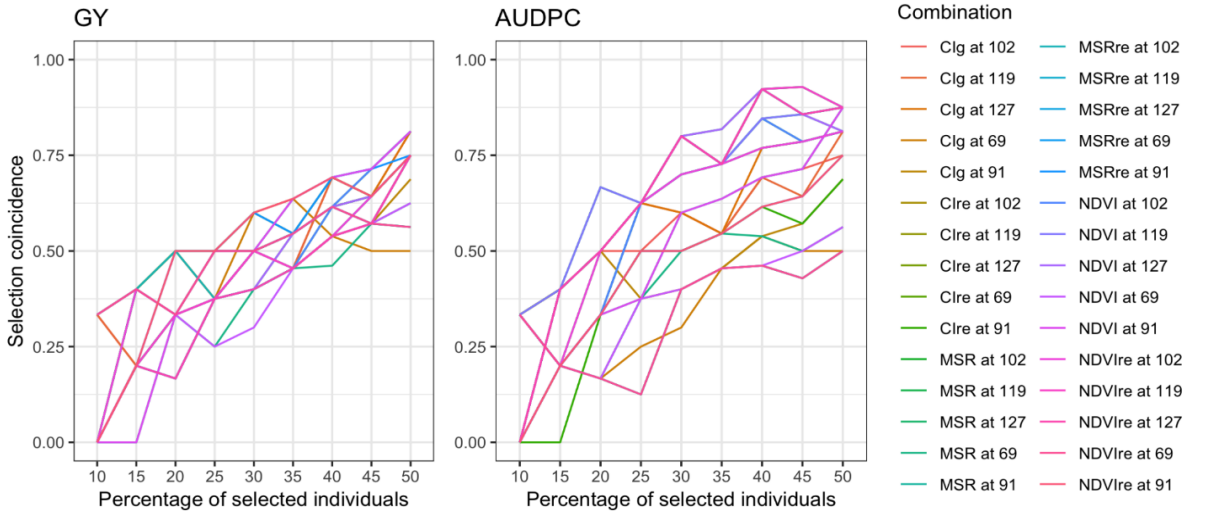

Figure S4. Coincidence of selection based on adjusted means of vegetation indices (CIg, MSR, NDVI, CIre, MSRre and NDVIre) at five flight dates (69, 91, 102, 119 and 127 days after sowing) with grain yield (GY) and anthracnose area under disease progress curve (AUDPC) of sorghum field in College Station, TX. Individuals with highest GY and lowest AUDPC values were selected.

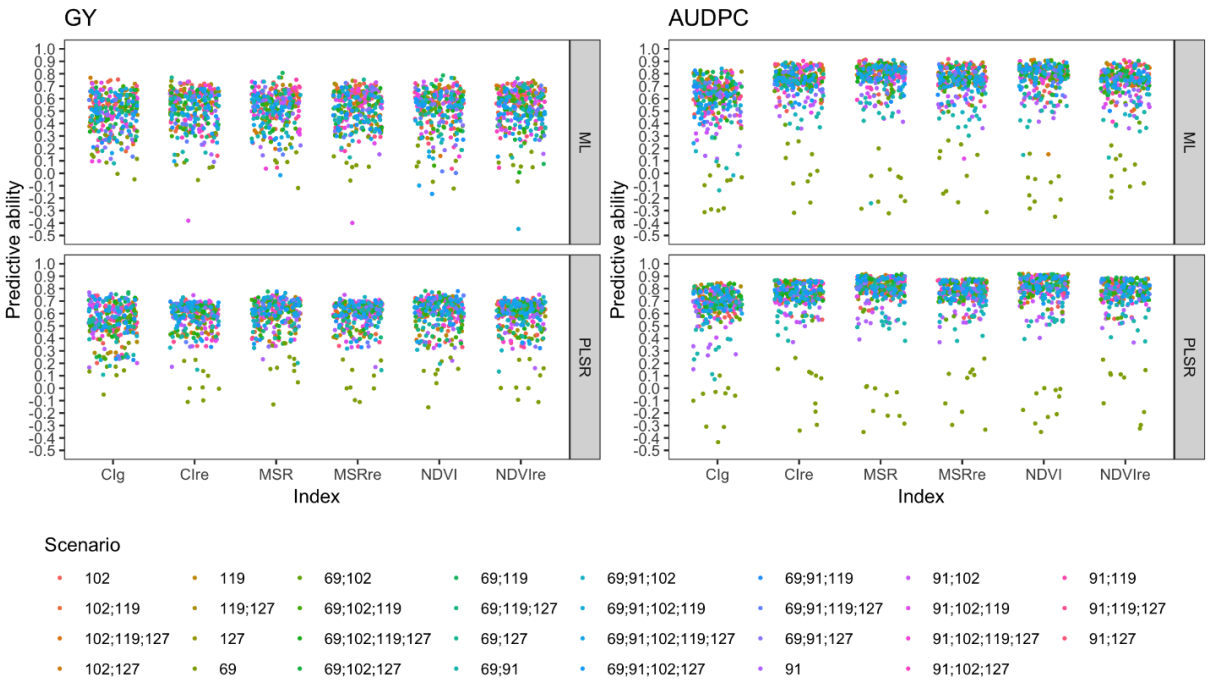

Figure S5. Three-entry (method × index × scenario) scatter plot of predictive ability of grain yield (GY) and area under disease progress curve (AUDPC) plot values of sorghum hybrids in College Station, TX. Scenarios are flight dates (in DAS) and its combinations. Each point is a prediction ability estimated in the validation process. Hence, each scenario is represented by ten points in the graph.

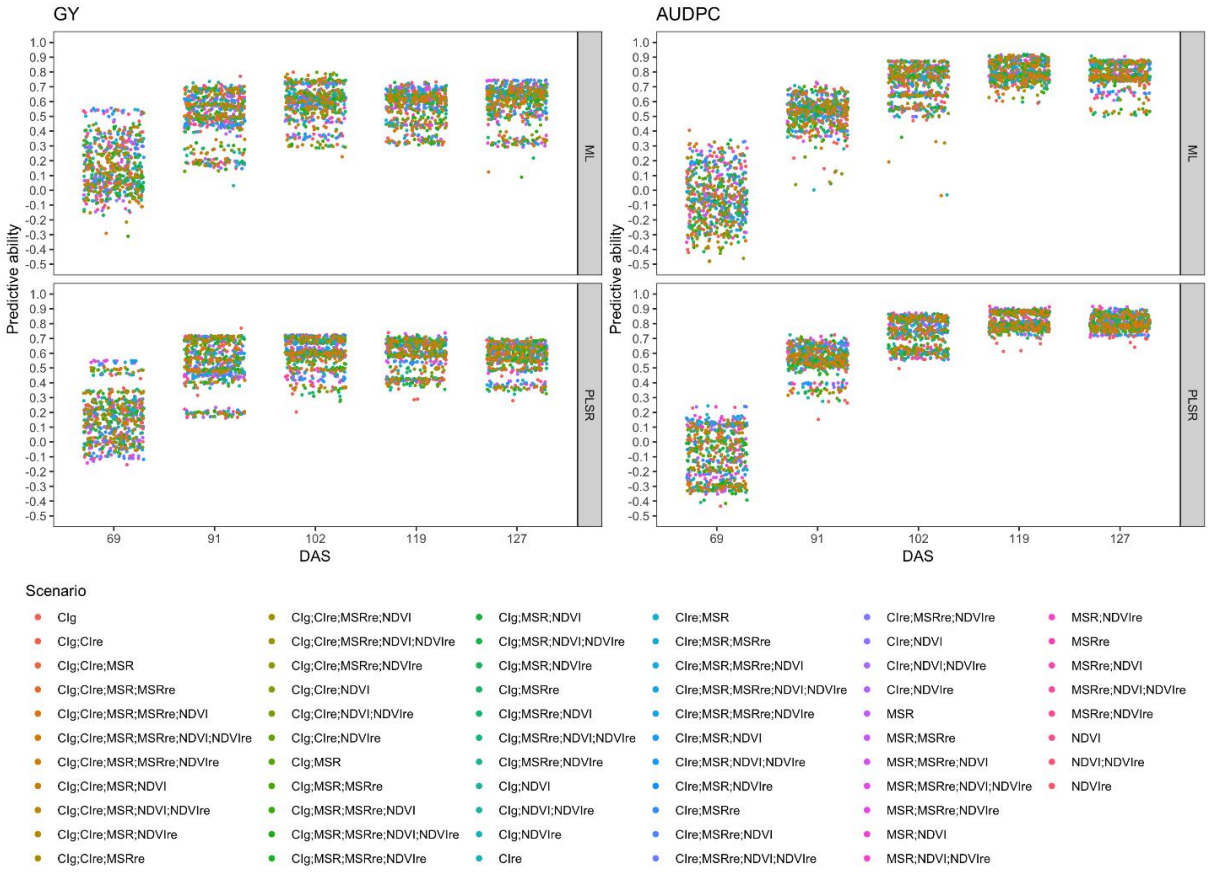

Figure S6. Three-entry (method × index × scenario) scatter plot of predictive ability of grain yield (GY) and area under disease progress curve (AUDPC) plot values of sorghum hybrids in College Station, TX. Scenarios are indices and its combinations. Each point is a prediction ability estimated in the validation process. Hence, each scenario is represented by ten points in the graph.

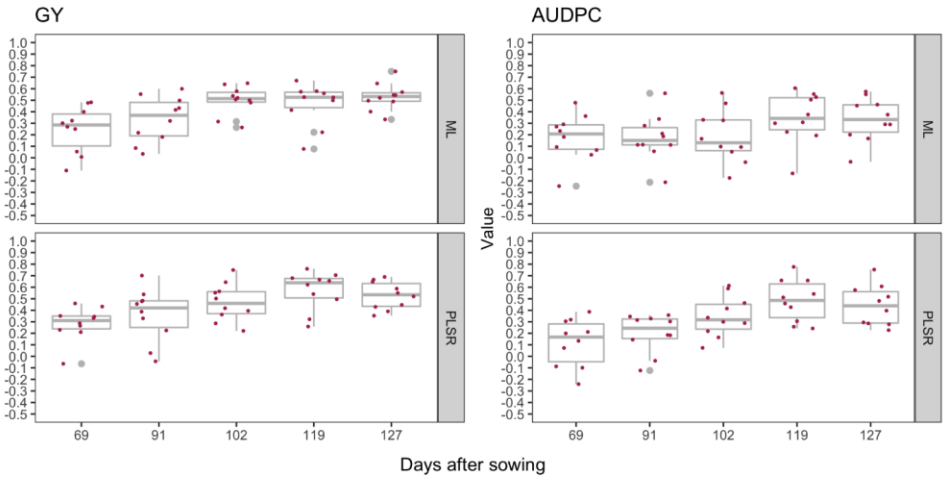

Figure S7. Predictive abilities of grain yield (GY) and area under disease progress curve (AUDPC) using the combination of four spectral bands (Green, Red, Red edge, Near infra-red) from five flights (69, 91, 102, 119, and 127 days after sowing) over sorghum hybrids in College Station, TX.

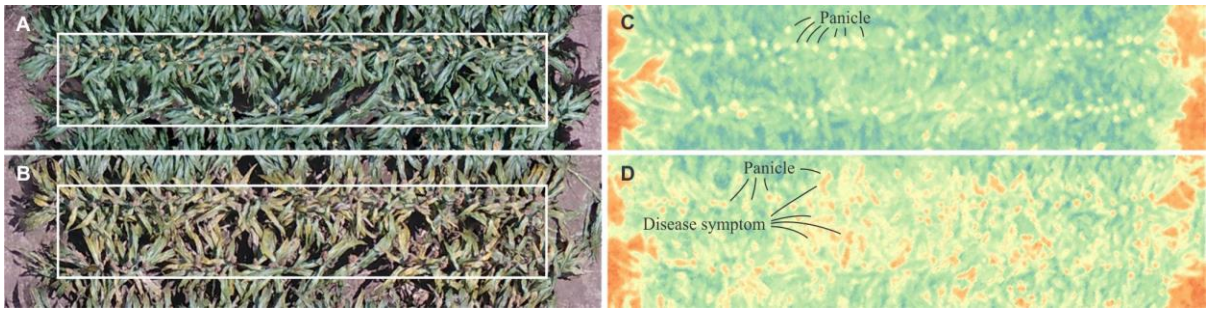

Figure S8. A: Sorghum plot RGB image at 91 days after flowering (DAS); **B**: Sorghum plot RGB at 119 DAS. On A and B, the white rectangle represents plot boundaries; **C**: Sorghum plot pseudo color NDVI at 91 DAS; **D**: Sorghum plot NDVI at 119 DAS.

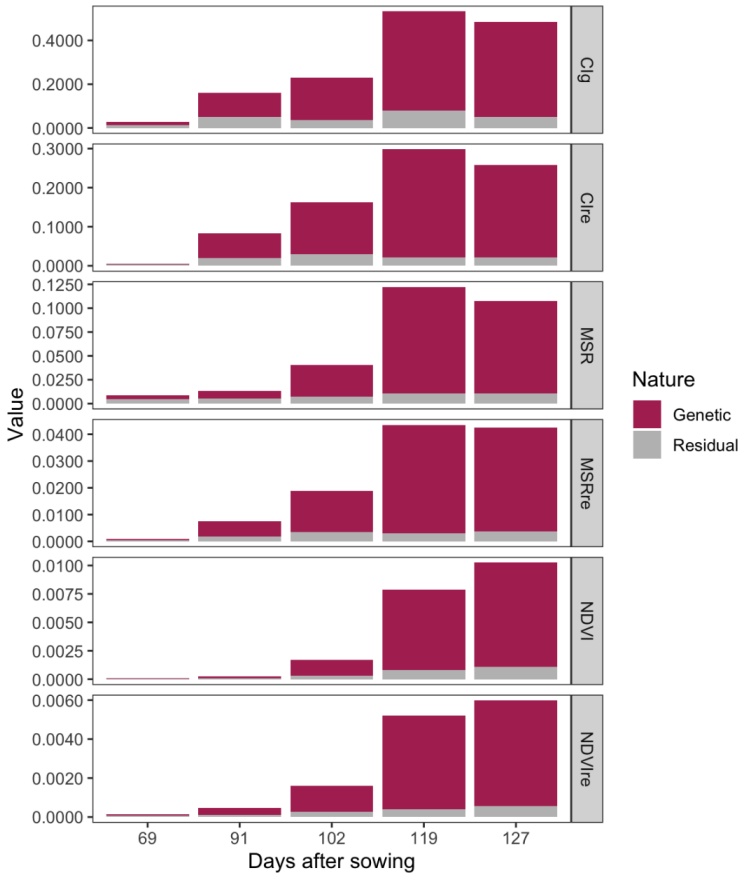

Figure S9. Progress of genotypic and residual variances on five flight dates for chlorophyll index (CIg), modified simple ratio (MSR), normalized difference vegetation index (NDVI), and their formulas based on the red edge wavelength (CIre, MSRre and NDVIre, respectively) in sorghum hybrids in College Station, TX.

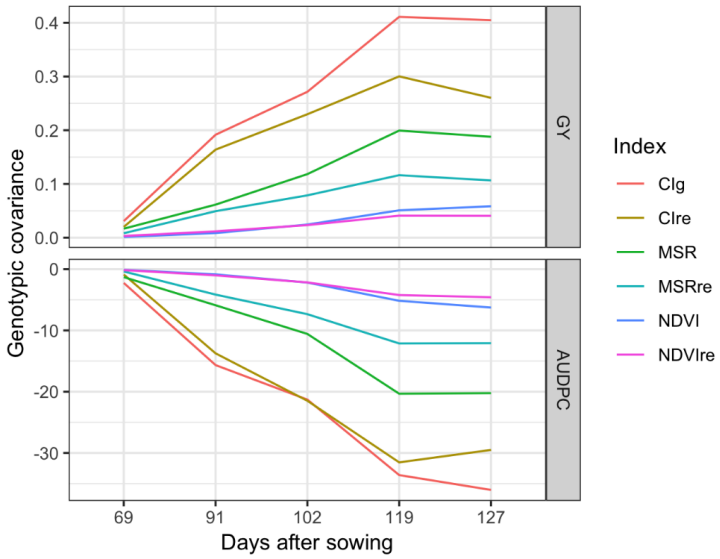

Figure S10. Progress of non-standardized genotypic covariances between main (GY and AUDPC) and secondary traits (CIg, MSR, NDVI, CIre, MSRre and NDVIre) in five flight dates (69, 91, 102, 119 and 127 days after sowing) over sorghum hybrids in College Station, TX.

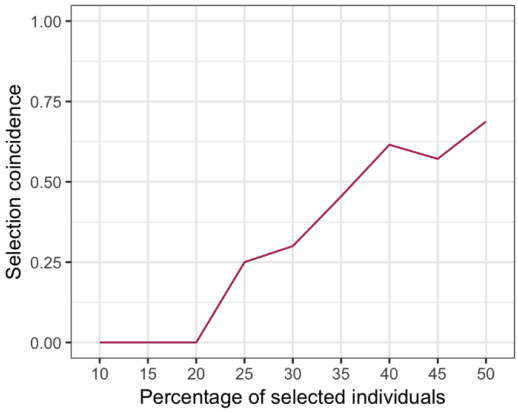

Figure S11. Coincidence of selection based on adjusted means of grain yield (GY) and anthracnose area under disease progress curve (AUDPC) of sorghum field in College Station, TX. Individuals with highest GY and lowest AUDPC values were selected.

3. A NOVEL WAY TO VALIDATE UAS-BASED HIGH-THROUGHPUT PHENOTYPING PROTOCOLS USING *IN SILICO* EXPERIMENTS FOR PLANT BREEDING PURPOSES

ABSTRACT

Protocol validation has been a key challenge to the establishment of high-throughput phenotyping (HTP) in breeding programs. We add to this matter by proposing an innovative way for designing and validating aerial imagery-based HTP approaches with *in silico* 3D experiments for plant breeding purposes. The algorithm is constructed following a pipeline composed of the simulation of phenotypic values, tridimensional modeling of trials, and image rendering. Our tool is exemplified by testing a set of experimental setups that are of interest in the context of plant breeding using a comprehensive case study. We report on how the choice of (percentile of) points in dense clouds, the experimental repeatability (heritability), the treatment variance (genetic variability), and the flight altitude affect the accuracy of HTP plant height estimation based on conventional structure-from-motion (SfM) pipelines. The evaluation of both the algorithm and the case study was driven by comparisons of the computer-simulated and the HTP-estimated values using correlations, regressions, and similarity indices. Our results showed that the 3D experiments can be adequately reconstructed by SfM, enabling inference-making. Moreover, it suggests that treatment variance, repeatability, and the choice of the percentile of points are highly influential over the accuracy of HTP. Conversely, flight altitude influenced the quality of reconstruction but not the accuracy of plant height estimation. Therefore, we believe that our tool can be of high value, enabling the promotion of new insights and further understanding of the events underlying the practice of high-throughput phenotyping.

Keywords: phenomics; simulation; structure-from-motion; aerial imagery

3.1. INTRODUCTION

High-throughput phenotyping (HTP) has been consistently drawing the attention of the scientific community as a field for potentially improving the rates of genetic gain (White et al., 2012; Araus and Cairns, 2014; Araus et al., 2018). A wide range of ground and aerial-based systems have been implemented to perform phenotyping, but the use of unmanned aerial systems (UAS) has stood out due to the relatively low cost, the ability to carry out large-scale in-field screenings, and the high spatial/temporal resolutions (Afonnikov et al., 2016). UAS are usually equipped with sensors that register the reflectance on specific bands for making inferences about the most diverse plant characteristics (White et al., 2012; Bowman et al., 2015; Araus et al., 2018).

The reflectance of canopy and soil is captured as digital images, which can be used in its raw form, to calculate point clouds, and/or stitched into mosaics. Evidences show that this type of data has potential for the measurement of traits such as plant height (Shi et al., 2016a; Pugh et al., 2018b; Malambo et al., 2018), lodging (Chu et al., 2017), grain yield (Edlich-Muth et al., 2016; Rutkoski et al., 2016), biotic stress (Stanton et al., 2017; Pugh et al., 2018a), biomass (Li and Shi, 2018; Wang et al., 2019), nitrogen status (Zaman-Allah et al., 2015; Tattaris et al., 2016), growth dynamics (Pugh et al., 2018b), and stand count (Varela et al., 2018). Most measurements can be performed with consumer-grade equipment, such as a quadcopter, an imaging sensor, and a processing algorithm. That is, it is

possible to improve phenotyping in the whole realm of agronomically relevant traits with accessible methodologies (Araus et al., 2018).

A key challenge to the application of HTP in plant breeding is to design and validate procedures for specific situations. For instance, the quality of phenotyping for a given trait, say maize plant height, may be influenced by a series of features such as flight altitude, image overlapping, experimental quality, genetic variation of the population, and the choice of (percentile of) points in a dense cloud that defines the height. It raises many questions as to how accurate HTP is expected to be and how it should be carried out in a given scenario. Nevertheless, an adequate validation with transferable results requires the test of HTP pipelines to be performed under consistent and representative conditions with minimization of errors.

Several factors have been reported to cause inaccuracies in the results of UAS-based HTP, which can be grouped into three main sources: UAS survey, data processing, and crop phenology (Grenzdörffer, 2014). Moreover, the nature of these elements can be environmental, such as wind and light conditions during flight (Mesas-Carrascosa et al., 2015; Tattaris et al., 2016; Chang et al., 2017); and technical, such as equipment limitations (Mesas-Carrascosa et al., 2015), processing algorithm errors (Grenzdörffer, 2014), and experimental error. Hence, having sufficiently controlled and representative experiments might be difficult. Nevertheless, these conditions are necessary to enable inferencing about the applicability and accuracy of HTP.

When conditions are not suitable for testing a given hypothesis, there is the possibility of generating *in silico* models with full control of the factors affecting the results. In this sense, simulation of artificial images can generate databases to overcome limitations of real datasets (Barth et al., 2018; Müller et al., 2018). Three-dimensional (3D) computer-generated scenes have been extensively applied for training models focused on several computer vision tasks such as semantic segmentation, feature detection, object tracking, autonomous navigation, and scene understanding (Di Cicco et al., 2017; Johnson-Roberson et al., 2017; Souza et al., 2017; Müller et al., 2018). These efforts include artificial modeling and plant image generation of agricultural species (Pradal et al., 2015; Di Cicco et al., 2017; Barth et al., 2018). Translating the use of simulations into the HTP context, this strategy could be implemented to generate fully controlled experiments and improve the reliability of estimates.

Using simulations for HTP would open up the possibility of testing new methods and experimental setups. It could be used to measure the potential of HTP under specific conditions and compare it to benchmark methodologies, adjusting the expectations. It could also be implemented for equipment dimensioning, which may lead to substantial experimentation savings. However, to the best of our knowledge, there is still no report on the possibility of using simulations for HTP validation using UAS-like data. In this sense, we aimed to propose and exemplify an innovative way to design and validate HTP methodologies with computer-generated 3D experiments. Precisely, this study was carried out to verify if reconstruction algorithms (e.g., structure-from-motion; SfM) can adequately reproduce synthetic experiments based on UAS-like imagery of crops; and present a comprehensive simulation case study testing the effect of the choice (percentile of) points that define estimated height in a dense cloud and experimental factors such as treatment variance (genotypic variance), experiment repeatability (heritability), and flight altitude on the accuracy of HTP of maize plant height.

3.2. MATERIAL AND METHODS

The "*Material and Methods*" section was divided into two major parts. In the first, we presented a new algorithm for building synthetic experiments, which was composed of phenotype simulation, 3D modeling of field trials, and image rendering (Figure 1). The code is available at github.com/giovannigalli/UASmachine. In the second, we tested the algorithm by applying it to a case study of plant height measurement in maize. This step was composed of image analysis and the comparison of several experimental setups obtained with variations of the simulation parameters.

3.2.1. The algorithm

3.2.1.1. Phenotype simulation

UAS-based high-throughput phenotyping is idealized to efficiently measure traits of a large number of treatments (genotypes) (Araus et al., 2018). This requires the use of incomplete block designs to properly accommodate all treatments and still adequately control experimental noise (uncontrolled random variations; Fritsche-Neto and Borém, 2015). In this sense, the following algorithm was written to simulate experiments following augmented block (common and increased treatments) and lattice designs. Thus, for simulation purposes, the treatment randomizations are obtained with *design.lattice* and *design.dau* functions from R's *agricolae* package (De Mendiburu, 2016).

Firstly, the experiment is set by choosing the experimental design, number of replications, number of treatments (regular and increased for augmented blocks), number of plants and rows per plot, the incomplete block field layout, overall mean, repeatability and treatment (genotypic) variance of the trait of interest, block variance, and sampling seed (Figure 1). Then, the synthetic phenotypes ($\boldsymbol{y}$) are obtained following the model:

$$\boldsymbol{m} + U\boldsymbol{l} + W\boldsymbol{b} + Z\boldsymbol{g} + \boldsymbol{\epsilon} = \boldsymbol{y}$$

where $\boldsymbol{m}$ is the overall mean; $\boldsymbol{l}$ is the replication effect [$\boldsymbol{l} \sim \mathcal{U}(-r, r)$, being r an scalar]; $\boldsymbol{b}$ is the incomplete block effect [$\boldsymbol{b} \sim \mathcal{N}(\boldsymbol{0}, I\sigma_b^2)$]; $\boldsymbol{g}$ is the treatment (genotype) effect [$\boldsymbol{g} \sim \mathcal{N}(\boldsymbol{0}, I\sigma_g^2)$]; and $\boldsymbol{\epsilon}$ is the error [$\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, I\sigma_\epsilon^2)$]. U, W and Z are the incidence matrices of replication, incomplete block and treatment, respectively. This protocol enabled the simulation of the phenotypic value of any normally distributed trait of interest to be used for further *in silico* 3D modeling.

Three output files are obtained from the phenotype simulation step. The first contains the identifications of plant, incomplete block, treatment, replication, and plant number in the plot. It also has the simulated phenotype of each plant, incomplete block effects, treatment effects, replication effects, and random errors. For further reference and analysis, the spatial indices (row and column) are also recorded. It also outputs a simplified file that contains the means by plot. In this file, plant identification is replaced by plot identification. Additionally, the function writes a file with the input parameters used in the 3D modeling step.

3.2.1.2. Modeling *in silico* field experiments

The phenotypes are translated into 3D scenes using Python's core functions and the Blender's API (bpy). It is assembled by setting the environment and positioning objects that compose the experiment. This procedure requires selecting a set of arguments and objects such as camera altitude (flight altitude), 3D plant models, 3D objects to represent ground control points, image overlapping, camera parameters, spacing between rows and plants, soil altitude, experiment and ground control point layout, and whether or not borders should be added (Figure 1). Other relevant inputs are imported from R exported files obtained in the phenotype simulation step. The incomplete block layout is manually set by informing the relative position of each block in the $[x, y]$ Cartesian space, e.g., for an experiment with two (1×2) incomplete blocks, it could be set as $[-0.5,0], [0.5,0]$.

The environment is constructed with the following characteristics:

- Lightning is set by using a 360° environmental texture captured with High Dynamic Range Imaging (HDRI; www.hdrihaven.com), with the main source of light (sun) set to be at nadir. The texture has no clouds to represent a clear sky condition;
- No rotation is applied to the texture, so the main source of light is still and constant across the entire rendering process;
- No motion is added to the objects to simulate the total absence of wind, technically not affecting the ability of the reconstruction software to recreate the scene objects;
- The soil surface is drawn as a plane with the procedural texture generated with Blender shaders;
- The plants are accordingly positioned in the scene, given the incomplete block layout and spacing between plants and rows from the experimental virtually designed in the previous section.

The *in silico* experiments can be composed of any three-dimensional plant models the algorithm is fed. Each plant in the scene is a copy of a randomly chosen basic model with some controlled modifications. The plants are then assigned the corresponding trait value simulated in R. Furthermore, plants are added as border on the north and south experiment limits with $\boldsymbol{m} + \boldsymbol{e}$ phenotype [$\mathcal{N}(\boldsymbol{m}, \sigma_\epsilon^2)$].

To avoid eventual positioning issues, coded ground control points (GCPs) generated in Agisoft Metashape Professional version 1.5.1 (build 7618; 64 bit) using the *print markers* tool are imported into the scene. Hence, GCPs are positioned just outside the experiment boundaries based on the GCP layout provided by the user.

3.2.1.3. Image rendering and flight plan

Image rendering is the estimation of the pixel value in an image according to the scene objects, environment and camera settings (Figure 1). Thus, the camera features are set to simulate a specific imaging sensor. It includes image resolution (vertical and horizontal), focal length, sensor height, and sensor width. The inherent camera pixel size is estimated for further image analysis (Tulsiani et al., 2018) as:

$$P_{vertical} = 1/\left(\frac{vertical\ resolution}{sensor\ height}\right)$$

and

$$P_{horizontal} = 1/\left(\frac{horizontal\ resolution}{sensor\ width}\right)$$

The 'flight plan' (set of positions where images should be taken in the scene) is set based on the distances between images given that vertical and horizontal angles of view (AOV; Mikhail et al., 2001) are:

$$AOV_{vertical} = 2 \times \arctan\left(\frac{sensor\ height}{2 \times focal\ length}\right)$$

and

$$AOV_{horizontal} = 2 \times \arctan\left(\frac{sensor\ width}{2 \times focal\ length}\right)$$

Hence, lateral (D_x) and frontal (D_y) distances between the center of images are calculated as:

$$D_{frontal} = 2 \times h \times \tan\frac{AOV_{vertical}}{2} \times (1 - overlap)$$

and

$$D_{lateral} = 2 \times h \times \tan\frac{AOV_{horizontal}}{2} \times (1 - overlap)$$

where h represents the flight altitude. Local coordinates (m) of the camera (known without error) are recorded (label, easting (x), northing (y), and altitude (z)) to be used as EXIF metadata.

The rendering process (calculation of pixel values) is done using Blender's Cycles engine with CPU and/or GPU. The outputs of this procedure are UAS-like computer-generated images, which are written in TIFF format with no compression. In this procedure, a file with local coordinates of the images (EXIF metadata), a file with local coordinates of every plant, a file with GCP local coordinates, and a file containing pixel size and camera focal length are also generated.

3.2.2. Case study: plant height in maize breeding

3.2.2.1. Experimental setups

To exemplify the usefulness of this novel approach, we simulated experiments in which plant height (PH_{sim}) is the trait of interest. With the intent of assessing the ability to estimate plant height based on a conventional HTP SfM pipeline under different setups, 54 scenarios were tested by combining two levels of treatment variance (σ_g^2), three experimental repeatabilities (R^2), three flight altitudes (h), and three sampling seeds.

The phenotype simulation step was performed with the following parameters: three sampling seeds were utilized so the simulations are repeatable. The simulated experiments were designed in lattice with two replications, 14 blocks, and 49 genotypes. The incomplete block layout used in this study was 7×2 (Figure S1). Each incomplete block had seven single-row plots with 12 plants. The parameters σ_g^2 (genotypic variance) and $R^2 = \sigma_g^2/(\sigma_g^2 + \sigma_\epsilon^2)$ (repeatability/heritability) were controlled to generate scenarios that could be observed in certain phases of a breeding program. In this sense, σ_g^2 was assumed to be 0.0005 and 0.05, producing a range of genotypic values (between the smallest and tallest treatments) of about 0.10 and 1.20 m, respectively. These values represented populations of low and high genotypic variation. Also, three levels of repeatability were considered, namely 0.3, 0.6 and 0.9, with the intent of simulating contrasting levels of experimental quality within each level of genotypic variation. Additionally, the experimental factor of block was added with variance (σ_b^2) equal to 0.0005, and the replication effect ($\boldsymbol{l}$) with limits of ± 0.05. The overall mean ($\boldsymbol{m}$) was set to 2 m.

The choice of parameters resulted in phenotypic values comprised within the expected range for maize plant height in field experiments. Mean of minimum, overall mean, and mean of maximum values of PH_{sim} across scenarios were 1.40, 2.00, and 2.58 m when σ_g^2 = 0.05 and 1.92, 1.99, 2.07 m, when σ_g^2 = 0.0005 (Figure S2). That is, for σ_g^2 = 0.05 there was a phenotype mean deviation of ± 0.60 m and of ± 0.08 for σ_g^2 = 0.0005. After the simulation of phenotypes, the scenes were generated with the steps and parameters described hereon.

The utilized plant models were eight hand-drawn 3D maize plants with realistic morphological structures (Figure 2). The models differed on the number of leaves (12 to 15), and ears (1 or 2), and in the structure of stalk and tassel. Images of leaf, stalk, tassel and ear were assigned as textures to the models. As the models were copied to compose the experiment, controlled modifications were applied using Blender's wave and deform modifiers to generate small random changes to the plant morphology. The last modification was a random rotation of the plant copies over their z axys. Additionally, to facilitate SfM reconstruction, the textures of each plant were mixed with random RGB colors. At last, in order to assign the appropriate plant height to each plant copy, Blender's resize function was utilized to change the extension of the objects in the z axis to the corresponding simulated phenotype. Therefore, each plant in the scene had unique visual/morphological features and the appropriately assigned phenotype.

As the plants were imported and modified, they were placed on the scene at a given coordinate following the 7×2 layout and the spacing between plants (0.33 m) and rows (0.5 m). Borders were added. The soil altitude was set at $z = 0$ to facilitate further analysis. In addition, a total of six ground control points were placed.

Camera parameters were set to emulate common and accessible equipment (e.g., Phantom 4 Pro imaging sensor). Therefore, images were calculated in 5472×3648 pixels with a focal length of 24 mm. Further, the sensor's height and width were 8.8 and 13.2 mm, respectively. The overlap (frontal and lateral) was 0.8 (80%), and the images were rendered at $h =$ 20, 40, and 60 m to simulate different spatial resolutions. The values of h resulted in the rendering of 91, 40, and 35 images, respectively.

Analysis were run in an Intel i9 processor (18 cores; 36 threads), 128 GB RAM desktop computer. For comparison purposes, a set of images (Figure S3) were rendered using two other processors, namely, Intel versions i5 and i7 (Table S1).

3.2.2.2. Image analysis

Image processing was performed in Agisoft Metashape Professional version 1.5.1. Camera features, such as focal length and pixel size were imported. The images were aligned, and their position optimized using the GCPs (automatically detected). The alignment was followed by digital point cloud estimation, calculation of digital elevation models (DEMs), and orthomosaics (settings and parameters can be found in Table S2). An orthomosaic was imported into QGIS software (version 3.0.2; QGIS Development Team, 2018) to be used as a reference for drawing plot boundaries with the *v.mkgrid* function of GRASS plugin (version 7.4.0; GRASS Development Team, 2018). To avoid interplot leaf overlapping, a negative buffer of 0.1 m was applied using the *Buffer* function. The shapefile of plots and the point clouds were imported into R environment (R Core Team, 2019) with the *lid*R (Roussel and Auty, 2019) and *raster* (Hijmans et al., 2015) libraries and segmented by $[x, y]$ plot boundaries.

Each point in the clouds was composed of a set of $[x, y, z]$ local coordinates. A one-measure-by-plot strategy was used to estimate point cloud plant height (PH_{htp}). Since precision georeferencing was performed with

GCPs, soil points were assured to be at $z = 0$ m. Therefore, soil height was not subtracted from the canopy height for PH_{htp} estimation. The points in each plot were split into two groups at the p-percentile of the ordered z axis values. The z value at the p-percentile was said to be the height of the plot. In this sense, plant height was obtained for each plot from the dense point cloud using the following methods:

1. Using a sequence of percentiles within a range of integer values. The PH_{htp} was estimated based on percentile values from 1 to 100;
2. Using an optimization function. The percentile that maximizes the Pearson's Product-Moment Correlation between PH_{htp} and PH_{sim} of all plots simultaneously was optimized with:

$$p_{opt} = \text{argmax} f(p)$$

   where function $f(p)$ correlates PH_{sim} and PH_{htp}, given that p_{opt} is the percentile that makes $f(p_{opt})$ a local maximum. This was performed using R's *optim* function with initial value of $p = 0.9$, bound to a lower value of 0.8 and upper value of 1, and using the *L-BFGS-B* method (Byrd et al., 1995). Once p_{opt} was found for each simulation scenario, the image-based plant height was estimated from the point cloud. These were named optimized percentiles.
3. Using the percentile that maximizes the correlation between PH_{sim} and PH_{htp} for each scenario (obtained from method one). These were named optimal percentiles;
4. Using a simplified range of 80th, 85th, 90th, 95th, and 100th percentiles.

Further statistical analyses were implemented to assess the ability of the pipeline to reproduce the simulated phenotypic values and determine the influence of factors composing the scenarios, as described below.

3.2.2.3. Scenario comparison

The ability to faithfully reproduce simulated experiments was studied by regressing PH_{sim} on PH_{htp} for each scenario following:

$$\boldsymbol{PH_{htp}} = \boldsymbol{\beta} + \boldsymbol{PH_{sim}}\lambda + \boldsymbol{\epsilon}$$

where PH_{htp} is the point cloud-estimated plant height, PH_{sim} is the simulated phenotypic value fit as a covariate, $\boldsymbol{\beta}$ is the intercept, λ is the regression coefficient associated with PH_{sim}, and $\boldsymbol{\epsilon}$ is the residual $[\mathcal{N}(\mathbf{0}, \sigma_{\epsilon}^2)]$. The values of PH_{sim} and PH_{htp} were presented as deviations from the mean of PH_{sim}. Therefore, the intercept represents the shift of PH_{htp} in space (z axis) when compared to PH_{sim} (e.g., the mean height of plants 1 m higher than it should be). Hence, the best value would be zero. The regression coefficient – or bias – (λ) represents the shrinkage of the range of heights (e.g., if the value is 2, the point cloud spans twice the space of z it should). The ideal value for this parameter is 1.

To assess the indirect selection efficiency, the ability to select for plant height based on image-based plant height, the following procedure was carried out. The model presented under the "*Phenotype simulation*" section was run, having PH_{sim} and PH_{htp} as dependent variables for the estimation of the independent vector of genotypes ($\hat{\boldsymbol{g}}$). For this, mixed models having replicate as a fixed effect, incomplete block, and genotypes as random were fit in R's *lme4* package (Bates et al., 2015) for each scenario. The best linear unbiased predictions (BLUPs; $\hat{\boldsymbol{g}}$) of genotypes were obtained for PH_{sim} ($\hat{\boldsymbol{g}}_{sim}$) and PH_{htp} ($\hat{\boldsymbol{g}}_{htp}$) for selection comparisons. Assuming that the higher the value of $\hat{\boldsymbol{g}}$

the best the treatment, the values were ordered and the 10 superior treatments selected for each scenario. The selection similarity (J) between $\hat{\boldsymbol{g}}_{sim}$ and $\hat{\boldsymbol{g}}_{htp}$ was estimated using Jaccard's index:

$$J(\hat{\boldsymbol{g}}_{sim}, \hat{\boldsymbol{g}}_{htp}) = \frac{\hat{\boldsymbol{g}}_{sim} \cap \hat{\boldsymbol{g}}_{htp}}{\hat{\boldsymbol{g}}_{sim} \cup \hat{\boldsymbol{g}}_{htp}}$$

given that the top 10 treatments are assigned value 1 (selected) and the other are assigned value 0 (not selected).

To test the significance of the factors expected to affect the relationship between simulated and estimated plant height values, a fixed linear model was fit for six analysis that differed regarding the dependent variable ($\boldsymbol{\gamma}$) following:

$$\boldsymbol{\gamma} = \boldsymbol{\mu} + Si + G\boldsymbol{v} + H\boldsymbol{r} + F\boldsymbol{a} + \boldsymbol{\epsilon}$$

where $\boldsymbol{\gamma}$ was:

1. The percentile values that maximize the correlation between PH_{sim} and PH_{htp}. This analysis was performed to test the effect of main factors on the establishment of the percentile;
2. A vector of Fisher-Z transformed correlations between PH_{sim} and PH_{htp}. PH_{htp} values are those estimated from the previously mentioned methods ("*Image analysis*" section), except number one. That is, the tested percentiles were p = 80, 85, 90, 95, 100, optimal and optimized. In this case, the percentile was included as an independent model variable ($P\boldsymbol{c}$) to test its effect on the correlations;
3. A vector of Fisher-Z transformed correlations between PH_{sim} and PH_{htp} obtained with the 90th percentile. This is further explained in the "results" section;
4. The values of intercepts ($\widehat{\boldsymbol{\beta}}$) obtained from the regression of PH_{sim} on PH_{htp}. This analysis was done to test the effect of the main experimental factors on the position shift of PH_{htp} in space compared to PH_{sim}.
5. The values of the regression coefficient ($\hat{\lambda}$) obtained from the regression of PH_{sim} on PH_{htp}. This analysis was done to test the effect of the main experimental factors on the spatial shrinkage of PH_{htp} in space compared to PH_{sim}.
6. The Jaccard's index values (J) from the selection based on PH_{sim} and PH_{htp}, done to test the effect of the main experimental factors on the index.

$\boldsymbol{\mu}$ is the overall mean; i is the effect of seed; $\boldsymbol{v}$ is the effect of σ_g^2 (with two levels: 0.05 and 0.0005); $\boldsymbol{r}$ is the R^2 effect (with three levels: 0.3, 0.6, and 0.9); $\boldsymbol{a}$ is the effect of h (with three levels: 20, 40, and 60), ($\boldsymbol{c}$ is the effect of percentile), and $\boldsymbol{\epsilon}$ is the residual [$\mathcal{N}(\mathbf{0}, I\sigma_\epsilon^2)$]. S, G, H, (P), and F are the incidence matrices that relate the independent variables to $\boldsymbol{\gamma}$. In some circumstances, interactions of main factors were added to the model. Finally, factors identified as significant in the ANOVA had their means compared by Tukey test. Analysis 2 and 3 were performed using transformed data; however, results were presented with non-transformed correlation values.

3.3. RESULTS

3.3.1. Image rendering and experiment reconstruction

Among the main constraints of simulation studies are hardware requirement and computational time, demanding feasibility tests. In this sense, our results showed that the time required to render an image of the experiment scene (Figure S3) was highly dependent on the hardware (Figure 3). The computer with an i5 processor took nearly up to 10 times longer than an i9 processor. Additionally, resolution implies an exponential increase in rendering time. Furthermore, flight altitude seems to have a greater effect on rendering time on less powerful processors.

Each flight at 20 m generated a total of 6.76 Gb of images covering an area of approximately 590 m^2; the flight at 40 m generated 2.97 Gb, representing an area of ~1,330 m^2; and the flight at 60 m a total of 2.60 Gb with an area of ~2,740 m^2. These values represent ratios of 11.7, 2.29, and 0.97 Mb m^{-2} for flights at 20, 40, and 60 m.

Once the rendering process was complete, we verified that the photogrammetry software Metashape was able to adequately align the simulated images and reproduce the synthetic experiments as orthomosaics, DSMs, and point clouds (Figure S4). The survey statistics for all simulations are present in supplemental Table S3. The estimated flight altitude error, the difference between real camera position in z and its estimated average height above ground level, ranged from 0.4 to 8.6 with mean 2.58 m across scenarios. The mean difference was higher for the 20 m flights (3.47), followed by the ones at 60 (2.01) and 40 (2.01). The ground resolution, the effective resolution of all images, had a mean of 2.00, 4.01, and 6.01 mm px^{-1} for the images taken at 20, 40, and 60 m of altitude, respectively. The reprojection error, the distance between the original and reconstructed projection (error root mean square across all images), had a range of 0.42 to 1.12 and a mean of 0.72 pixels. Overall, images rendered with the camera at 20 m presented higher mean error (1.06) when compared to those taken at 40 m (0.53) and 60 m (0.57). The GCP location error estimate, root mean square error for estimated and real GCP position presented values as low as 0.01 cm and as high as 7.61 cm, with increasing precision at higher flight altitudes. The aforementioned parameters were neither influenced by treatment variance nor by repeatability. For more information on survey parameters refer to the Agisoft Metashape user manual.

A representation of the 3D reconstruction of a set of plots is shown in Figure 4. A visible interference of the flight altitude could be noticed in the shape of the point clouds, where the lower flight (A: 20 m) resulted in a visible sharper reconstruction of the plants when compared to the other altitudes (B: 40; and C: 60 m). The effect of experimental repeatability (D and G: 0.3; E and H: 0.6; and F and I: 0.9) was easier to visually notice when treatment variance was high (D; E; and F) due to the higher magnitude of the simulated errors. The effect of these factors was tested with the appropriate statistical approaches and is presented in the following sections (Table 1, Table 2).

3.3.2. The effect of percentile value on the correlations

The percentile of points utilized to estimate plant height presented a strong influence on the correlation value between PH_{htp} and PH_{sim} (Figure 5). Each gray line represents a simulated scenario and the colored lines (local regression) present the tendency of each tested level for the controlled experimental factors (σ_g^2, R^2, and h). Four main inferences could be drawn from this analysis. First, gray lines show that the best percentile is specific to each

scenario. Second, it could be observed that varying the percentile from one to 100 led to an irregular, linear-like, increasing trend on the correlation. However, from the 90 to the 100th percentile, the correlation tended to decrease. Third, the σ_g^2 levels seem to have great influence over the trends. For example, the mean correlation of 0.75 obtained with $\sigma_g^2 = 0.05$ and the low percentile of 30 was never reached when $\sigma_g^2 = 0.0005$. Fourth, as can be seen by the gray lines, whenever treatment variance was high, the changes in correlation due to percentile change were more consistent across scenarios. Finally, the levels of R^2 and h, however, did not show a clear distinction over the percentile parametric space.

Varying the percentile, it was possible to identify the value that leads to the highest possible correlation between PH_{htp} and PH_{sim} for each simulation. The ANOVA of the optimal percentile showed the significant statistical influence of treatment variance and repeatability over its determination (Table 1 Analysis 1, Figure S5). Accordingly, Tukey test showed contrasting groups of means for σ_g^2 (0.0005 = 91.04a; 0.05 = 87.04b) and R^2 (0.9 = 90.56a; 0.6 = 90.11a; 0.3 = 86.44b) (Table 2 Analysis 1). Additionally, comparing the optimal percentile with the one obtained with the optimization function, it could be seen that the values matched in four out of 54 simulations.

ANOVA revealed the influence of the percentile (summarized to 80, 85, 90, 95, 100, optimal, and optimized) over the establishment of the correlations (Table 1 Analysis 2). Tukey test showed that the optimal and the 90th percentile correlation means did not statistically differ (0.85a and 0.84a, respectively; Table S4). The optimized percentile (0.78c) only outperformed the 100th (0.67d). The other percentiles had intermediate performance (85 = 0.83ab; 95 = 0.82ab; and 80 = 0.78c). The ANOVA also revealed a significant interaction between percentile levels and σ_g^2 (Table 1 Analysis 2). Performing the Tukey test under each level of σ_g^2, the optimal and the 90th percentile presented similar values (0.73a and 0.72ab for σ_g^2 = 0.0005; 0.97a and 0.97ab for σ_g^2 = 0.05, respectively).

3.3.3. The effect of σ_g^2, R^2, and h on the correlations

In order to simplify the test of the simulation factors, the correlations obtained with the 90th percentile were utilized (Figure 6). The effects of seed, treatment variance, repeatability, and the interaction between treatment variance and repeatability were significant (Table 1 Analysis 3). According to the Tukey test (Table 2 Analysis 3), when treatment variance was equal to 0.05, the correlation (0.97a) was higher than when this parameter was 0.0005 (0.72b). Regarding repeatability, 0.9 (0.86a) led to higher correlations than 0.6 (0.85b) and 0.3 (0.82c). However, the means of correlations under different altitudes did not statistically differ.

3.3.4. The effect of σ_g^2, R^2, and h on the regression of PH_{sim} on PH_{htp}

The average value of the intercepts ($\hat{\boldsymbol{\beta}}$) was -0.43, ranging from -0.72 to 0.11. Significant statistical influence was revealed for h in the ANOVA (Table 1 Analysis 4). According to Tukey test, all contrasts were significant with increasing shift as the flight altitude increased (20 = -0.26a; 40 = -0.47b; and 60 = -0.57c; Table 2 Analysis 4).

Point cloud shrinkage was estimated as the regression coefficients of PH_{sim} on PH_{htp}. The average value was equal to 0.92, varying from 0.71 to 1.26. From ANOVA, it was possible to detect a significant effect of h and σ_g^2

over the shrinkage of PH_{htp} when compared to PH_{sim} (Table 1 Analysis 5). Thus, the flight simulated at 20 m was less biased (1.01a) when compared to those at 40 (0.88b) and 60 m (0.86b) (Table 2 Analysis 5). Regarding the treatment effect on the shrinkage, the simulations with lower treatment variance were less shrunken (0.0005 = 0.96a and 0.05 = 0.88b).

3.3.5. The effect of σ_g^2, R^2, and h on the selection coincidence

Jaccard similarity index for selection based on PH_{htp} and PH_{sim} varied from 0.25 to 1.00 with a mean of 0.62 across scenarios. The ANOVA showed that treatment variance is the only factor affecting the index (Table 1 Analysis 6). In this sense, the scenarios where σ_g^2 was 0.05 presented higher index means (0.81a) when compared to when σ_g^2 was 0.0005 (0.42b) (Table 2 Analysis 6).

3.4. DISCUSSION

3.4.1. Computer-generated experiments enable HTP validation

The results showed that it is possible to satisfactorily reconstruct computer-generated experiments based on UAV-like imagery in the context of HTP conventional pipelines. This statement is supported by the survey statistics (Table S3), which represents the ability of the SfM software to project images onto 3D structures. It is also supported by the high similarity of simulated and estimated plant height, especially when conditions were favorable. For instance, when treatment variance and experimental repeatability were high (0.05 and 0.9, respectively), the mean correlations between PH_{sim} and PH_{htp} was as high as 0.99. Next, the effect of each experimental factor on the quality of reconstruction and inferences driving HTP accuracy will be discussed.

3.4.2. Effect of percentile on the accuracy of HTP

The percentile of points used to estimate PH_{htp} showed strong influence over the accuracy of HTP (Figure 5). Overall, treatment variation and experimental repeatability presented significant influence over the value of percentile that maximizes the correlation between PH_{sim} and PH_{htp} (Table 1 Analysis 1). As a consequence, the value was shown to be highly specific to the evaluation scenario (Figure 5). In an attempt to automate the process of finding the best percentile, an optimization function was utilized. However, it led to the correct value in only 7.5% of the simulations. As Figure 5 shows, the line drawn from values of correlation as a function of percentile was irregular and composed by several local maxima impairing convergence to global maxima, especially under low treatment variance. For instance, the use of optimization functions to find the optimal percentile still needs improvement.

Although there was a percentile that maximized the correlation between simulated and estimated plant heights for each scenario, no significant difference between the correlations obtained with optimal and the 90th percentile was found (Table S4). Therefore, this can be regarded as a good overall fixed percentile for the simulated experiments. The effect of percentile on the accuracy of point cloud-based HTP measurement of plant height has

already been reported for maize (Li et al., 2016; Pugh et al., 2018b; Malambo et al., 2018). Li et al. (2016) showed that the 90th outperforms the 50th percentile. Malambo et al. (2018) presented result that suggests the 99th as superior to the 90, 95, and 100th percentiles. Pugh et al. (2018b) tested the 95th versus 100th and observed better results with the first. Our results are consistent with what has been reported in the sense that higher values lead to better correlations to a certain point. That is, a "buffer" should be maintained to account for experimental (e.g., low repeatability) and point cloud reconstruction errors.

3.4.3. Effect of treatment variance on the accuracy of HTP

Treatment variance, or genotypic variance, presented significant effect over the establishment of the optimal percentile, the correlations between PH_{sim} and PH_{htp}, the shrinkage of the point cloud, and the Jaccard's similarity index for selected genotypes from PH_{sim} and PH_{htp} (Table 1). This factor was the most influential on the establishment of the correlations with an increase of 0.25 units by changing σ_g^2 from 0.0005 to 0.05 (using the 90th percentile; Table 2 Analysis 3). Considering that treatment variation did not affect reconstruction (e.g., reprojection error and GCP error; Table S3), the inherent SfM processing errors could be regarded as similar for both levels. It means that when treatment variance is low, these errors have a more substantial impact on trait value relative to the populational variation. Accordingly, when treatment variance is high, the phenotypical differences between plots are easier to distinguish, leading to higher correlations to the ground truth.

Treatment variance was the only simulation parameter that significantly affected selection coincidence with a mean increase of 0.39 from the low to the high value (Table 2 Analysis 6). Therefore, greater treatment variance increases not only the correlations but also the accuracy of indirect selection. This event has critical implications for the process of plant breeding. The simulated populations are comparable to the early generation and advanced genotypes, featuring high and low genotypic variation, respectively. In this sense, SfM-based field HTP of plant height in maize could be applied in both breeding phases, but one should expect much higher selection accuracy when treatment variance in higher. Similar inferences were made by Pugh et al. (2018b) for plant height in grain sorghum. At last, the presented findings suggest the amount of genotype variation as one of the most important experimental factors for the success of field-HTP under controlled experiments.

3.4.4. Effect of experiment repeatability on the accuracy of HTP

Experimental repeatability, or heritability, showed a statistically significant effect on the establishment of the optimal percentile and correlations between simulated and estimated plant height (Table 1). The analysis indicated that the greater the repeatability, the higher the similarity of estimated and simulated values (Table 2 Analysis 3), which results from the larger variation of inner plot plant height as repeatability decreases. However, when the analysis was performed by treatment level, no influence of repeatability was observed on the correlations for the low treatment variance population (Table S5). This is associated to the phenotype simulation step, where the error variance was calculated as a function of treatment variance and experimental repeatability, leading to larger errors added to phenotypes under high treatment variance. The greater inner plot error also explains the broader range of optimal percentiles under this treatment variance level (Figure S5). Therefore, conclusions regarding the

influence of experimental repeatability on the accuracy of HTP were highly dependent on the amount of experimental error.

3.4.5. Effect of flight altitude on the accuracy of HTP

Variables related to the structure of the point cloud, such as shrinkage and location shift, were significantly affected by flight altitude. The data showed that the lower the flight, the better the plant reconstruction. That is reasonable since the decrease in ground resolution, two to three times when compared to the flights at 40 and 60 m, led to a visible smoothing effect on the surface of the point cloud (Figure 4). Despite the better reconstruction, aligning images rendered at lower altitudes presented difficulties, possibly due to the fewer number of features present in each image, a frequent event even in real experiments. As a consequence, when flight altitude was 20 m, reprojection, GCP and estimated flight altitude errors were larger.

Although the reconstruction quality was affected by flight altitude, it did not significantly affect either the optimal percentile, the correlations between PH_{sim} and PH_{htp}, or the genotype selection coincidence. That is, increasing ground resolution, a considerably expensive action in terms of processing time and data volume, did not lead to higher overall HTP accuracy. That contradicts Shi et al. (2016b), who mentions that the lack of adequate resolution as a limiting factor over the correlation between estimated and measured plant height. However, this statement is reasonable, and we believe that this would happen in our study at higher, non-evaluated altitudes. Nonetheless, we showed that there is a possibility to decrease the ground resolution and maintain accuracy, but the limit is still to be elucidated.

3.4.6. Reconstruction of *in silico* images

Time benchmarking showed that rendering *in silico* experiments was highly dependent on computational power (Figure 3). For example, if it takes 100 images to cover the experimental area, processing time would be around 17 hours to render the experiment at 5472×3648 pixels using a high-end (i9) processor. On the other hand, a low-grade (i5) processor can take about 200 hours to simulate the same experiment. Nevertheless, there is the possibility of adjusting other parameters that affect rendering time to make it feasible, such as image resolution. Additionally, our results suggest that repeating the simulation analysis is important as indicated by the significant effect of sampling seed (Figure S2; Table 1 Analysis 2 and 3). Therefore, when dimensioning a simulation experiment, one should adjust their expectations based on hardware and time availability.

Inferences on the quality of reconstruction could be drawn from reconstruction survey statistics (Table S3). The reprojection errors were consistent with the expected for annual crops (Näsi et al., 2018; Fathipoor et al., 2019). In some cases, the value was slightly higher in the simulated experiments, which is probably because the removal of high error points, a recurrent activity, was not performed to enable the comparison between scenarios. The GCP location error estimate was in the expected range for agricultural experiments (Grenzdörffer, 2014; Moeckel et al., 2018; Näsi et al., 2018). Additionally, the mean estimated flight altitude error, difference from real to estimated camera position in z was 2.58 m, which was in the range of values obtained in our physical maize trial surveys. These results suggest that the simulation was adequately reconstructed and representative of physical experiments using Agisoft software.

From the engineering perspective, one might be interested in maximizing the similarity between features obtained from simulated and estimated 3D structures. PH_{htp} was on average 0.43 m shorter than simulated plants. This phenomenon has been reported to be associated with sparse canopy structures (Grenzdörffer, 2014). Besides, as flight altitude increased, the height difference also increased (Table 2 Analysis 4). The analysis showed a reduction (underestimation; shrinkage) in the range of plant heights, except for the flight at 20 m, which was nearly unbiased (0.01; Table 2 Analysis 5). The flights at 40 and 60 m presented similar shrinkages (~0.13), meaning that the feature controlling this size reduction is determined by a breakpoint on spatial resolution reached between 20 and 40 m. It might be associated with the lower ability to reconstruct tassels atop plants as ground resolution decreases (Figure 4), following Malambo et al. (2018), who suggested the occurrence of bias as a consequence of SfM reconstruction inaccuracies.

The occurrence of differences in mean height and shrinkage between simulated and estimated datasets has some implications for the interpretation of the results. For example, the plants evaluated in our study would be taller than their estimated values. Nevertheless, from a plant breeding perspective, having the proxy trait accurately representing the metrics of the real trait is important, but the ability to distinguish the available genotypes is key. In this sense, the Jaccard index showed that it is possible to reach high (complete) similarity of selection using an HTP methodology, especially when treatment variation was high. This reinforces the previously mentioned good quality of reconstruction.

3.4.7. Potential of the tool and further improvements

As previously mentioned, it was possible to reconstruct *in silico* generated images under the conventional HTP pipeline. We presented an introduction of an innovative way to draw inferences about HTP in a fully controlled environment. We believe that, overall, the tool is useful for representative studies and inference-making. However, there is still room for improvement. For instance, the quality of results from this project could be further enhanced by creating a database of 3D models reconstructed from physical plant and textures (e.g., leaf, tassel, and stalk textures) to represent, even better, material experiments and make inferences more precise.

In this work, we compared three experimental factors that supposedly affect the accuracy of HTP. Conversely, as the quality of simulations was adequate, we impute that the tool can be used for testing other setups. For example, by controlling the environment of the scene, one could test: the influence of the different wind speeds (adding random or controlled movements to the 3D models); the effect of clouds (adding objects with known transparency between the source of light and the plants); the impact of irregularities in the soil surface (changing the shape of the object that represents the soil); if collecting data during different times of the day changes the outcome (switching the position and angle of the light source); the influence of different weed population sizes (adding a controlled number of individuals from a other species).

Regarding experimental practices, data collection, and processing, one can test: if collecting images in different angles can affect reconstruction (changing the rotation of the camera); if the camera's GPS $[x, y, z]$ or gyroscope (yaw, pitch, and roll) errors affect reconstruction (adding random errors with known reasonable magnitude to camera positioning and rotation); the effect of different sensors at the same flight altitude (changing camera specifications); the most adequate overlapping for reconstruction (changing the spacing between images in the flight plan); whether plant and row spacing (population stand) affect reconstruction (changing the number of plants); if the number of rows per plot affects the results in species with high leaf overlapping (varying plot shape);

the effect of blur, common on fixed wing UAS, on image alignment and reconstruction (by adding blur to the rendering process); the adequate number of GCP for the appropriate reconstruction of a given area size (adjusting the GCP number and layout); the effect of SfM algorithm (using two or more software packages such as Agistoft Metashape and Pix4D).

In simulations, all plant features can be controlled. Broadly, the tool can also be used to: validate HTP measurement of other morphological traits such as plant biomass and leaf area index; asses the possibility of identification of diseases and flowering time (control affected area with real or procedural textures). Besides inferences, the tool could also be used for building pipelines for the prediction of real datasets by: automatic image annotation and augmentation for specific classification/regression tasks; training and validation of algorithms for counting plants or structures such as ears, panicles, and fruits based on machine learning and classical computer vision; etc.

The code utilized to implement the simulations have been made fully available and is open to the community (github.com/giovannigalli/UASmachine). The studies mentioned under the current section can be implemented with minor additions to the core algorithm, not requiring advanced coding skills. Thus, we expect that the availability of such a tool will create opportunities for researchers to increase the rate of learning, broaden the knowledge on the matter, and improve the field of high throughput phenotyping.

3.5. CONCLUSION

In this work, we presented a novel way to address questions and improve HTP based on simulations. Our results showed that it is possible to reconstruct computer-generated UAS-like 3D experiments adequately and draw inferences from the results. We exemplified the tool by testing the effect of experimental repeatability, treatment variance, flight altitude, and the choice of the percentile of points selected from the dense cloud on the accuracy of HTP plant height estimation. The results suggest that treatment variance (genetic variability) was the experimental factor affecting the accuracy of HTP the most. Besides, the study of the repeatability (heritability) showed that high experimental errors can also affect accuracy. Conversely, the range of utilized flight altitudes influenced reconstruction but not the accuracy of phenotyping. Additionally, we showed that the adequate choice of percentile for estimating point cloud-based plant height is a critical step in the HTP pipeline. At last, we believe that the presented tool can provide the community with new insights and a better understanding of the events underlying the use of high-throughput phenotyping.

REFERENCES

Afonnikov, D.A., M.A. Genaev, A. V. Doroshkov, E.G. Komyshev, and T.A. Pshenichnikova. 2016. Methods of high-throughput plant phenotyping for large-scale breeding and genetic experiments. Russ. J. Genet. 52(7): 688–701. doi: 10.1134/S1022795416070024.

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

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.

Barth, R., J. IJsselmuiden, J. Hemming, and E.J. Van Henten. 2018. Data synthesis methods for semantic segmentation in agriculture: A Capsicum annuum dataset. Comput. Electron. Agric. 144(December 2017): 284–296. doi: 10.1016/j.compag.2017.12.001.

Bates, D., M. Mächler, B. Bolker, and S. Walker. 2015. Fitting Linear Mixed-Effects Models Using {lme4}. J. Stat. Softw. 67(1): 1–48. doi: 10.18637/jss.v067.i01.

Bowman, B.C., J. Chen, J. Zhang, J. Wheeler, Y. Wang, et al. 2015. Evaluating grain yield in spring wheat with canopy spectral reflectance. Crop Sci. 55(5): 1881–1890. doi: 10.2135/cropsci2014.08.0533.

Byrd, R.H., P. Lu, J. Nocedal, and C. Zhu. 1995. A Limited Memory Algorithm for Bound Constrained Optimization. SIAM J. Sci. Comput. doi: 10.1137/0916069.

Chang, A., J. Jung, M.M. Maeda, and J. Landivar. 2017. Crop height monitoring with digital imagery from Unmanned Aerial System (UAS). Comput. Electron. Agric. 141: 232–237. doi: 10.1016/j.compag.2017.07.008.

Chu, T., M.J. Starek, M.J. Brewer, T. Masiane, and S.C. Murray. 2017. UAS imaging for automated crop lodging detection: a case study over an experimental maize field. Auton. Air Gr. Sens. Syst. Agric. Optim. Phenotyping II 10218(May 2017): 102180E. doi: 10.1117/12.2262812.

Di Cicco, M., C. Potena, G. Grisetti, and A. Pretto. 2017. Automatic model based dataset generation for fast and accurate crop and weeds detection. IEEE Int. Conf. Intell. Robot. Syst. 2017-Septe: 5188–5195. doi: 10.1109/IROS.2017.8206408.

Edlich-Muth, C., M.M. Muraya, T. Altmann, and J. Selbig. 2016. Phenomic prediction of maize hybrids. BioSystems 146: 102–109. doi: 10.1016/j.biosystems.2016.05.008.

Fathipoor, H., H. Arefi, R. Shah-Hosseini, and H. Moghadam. 2019. Corn forage yield prediction using unmanned aerial vehicle images at mid-season growth stage. J. Appl. Remote Sens. 13(03): 1. doi: 10.1117/1.jrs.13.034503.

Fritsche-Neto, R., and A. Borém. 2015. Phenomics: How next-generation phenotyping is revolutionizing plant breeding. Phenomics How Next-Generation Phenotyping is Revolutionizing Plant Breed.: 1–142. doi: 10.1007/978-3-319-13677-6.

GRASS Development Team. 2018. Geographic Resources Analysis Support System (GRASS) Software, Version 7.2.

Grenzdörffer, G.J. 2014. Crop height determination with UAS point clouds. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch. 40(1): 135–140. doi: 10.5194/isprsarchives-XL-1-135-2014.

Hijmans, R.J., J. v. Etten, J. Cheng, M. Mattiuzzi, M. Sumner, et al. 2015. Package ' raster .' https://cran.r-project.org/web/packages/raster/raster.pdf.

Johnson-Roberson, M., C. Barto, R. Mehta, S.N. Sridhar, K. Rosaen, et al. 2017. Driving in the Matrix: Can virtual worlds replace human-generated annotations for real world tasks? 2017 IEEE International Conference on Robotics and Automation (ICRA). IEEE. p. 746–753

Li, W., Z. Niu, H. Chen, D. Li, M. Wu, et al. 2016. Remote estimation of canopy height and aboveground biomass of maize using high-resolution stereo images from a low-cost unmanned aerial vehicle system. Ecol. Indic. 67: 637–648. doi: 10.1016/j.ecolind.2016.03.036.

Li, J., and Y. Shi. 2018. Elucidating Sorghum Biomass , Nitrogen and Chlorophyll Contents With Spectral and Morphological Traits Derived From Unmanned Aircraft System. 9(October): 1–12. doi: 10.3389/fpls.2018.01406.

Malambo, L., S.C. Popescu, S.C. Murray, E. Putman, N.A. Pugh, 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(June 2017): 31–42. doi: 10.1016/j.jag.2017.08.014.

De Mendiburu, F. 2016. Package 'agricolae.' Stat. Proced. Agric. Res.

Mesas-Carrascosa, F.J., J. Torres-Sánchez, I. Clavero-Rumbao, A. García-Ferrer, J.M. Peña, et al. 2015. Assessing optimal flight parameters for generating accurate multispectral orthomosaicks by uav to support site-specific crop management. Remote Sens. 7(10): 12793–12814. doi: 10.3390/rs71012793.

Mikhail, E.M., J.S. Bethel, and J.C. McGlone. 2001. Introduction to Modern Photogrammetry. John Wiley & Sons Inc., New York.

Moeckel, T., S. Dayananda, R.R. Nidamanuri, S. Nautiyal, N. Hanumaiah, et al. 2018. Estimation of vegetable crop parameter by multi-temporal UAV-borne images. Remote Sens. 10(5): 1–18. doi: 10.3390/rs10050805.

Müller, M., V. Casser, J. Lahoud, N. Smith, and B. Ghanem. 2018. Sim4CV: A Photo-Realistic Simulator for Computer Vision Applications. Int. J. Comput. Vis. 126(9): 902–919. doi: 10.1007/s11263-018-1073-7.

Näsi, R., N. Viljanen, J. Kaivosoja, K. Alhonoja, T. Hakala, et al. 2018. Estimating biomass and nitrogen amount of barley and grass using UAV and aircraft based spectral and photogrammetric 3D features. Remote Sens. 10(7): 1–32. doi: 10.3390/rs10071082.

Pradal, C., C. Fournier, P. Valduriez, and S. Cohen-Boulakia. 2015. OpenAlea: Scientific workflows combining data analysis and simulation. ACM Int. Conf. Proceeding Ser. 29-June-20. doi: 10.1145/2791347.2791365.

Pugh, N.A., X. Han, S.D. Collins, J.A. Thomasson, D. Cope, et al. 2018a. Estimation of plant health in a sorghum field infected with anthracnose using a fixed-wing unmanned aerial system. J. Crop Improv. 32(6): 861–877. doi: 10.1080/15427528.2018.1535462.

Pugh, N.A., D.W. Horne, S.C. Murray, G. Carvalho, L. Malambo, et al. 2018b. Temporal Estimates of Crop Growth in Sorghum and Maize Breeding Enabled by Unmanned Aerial Systems. tppj 1(1): 1–10. doi: 10.2135/tppj2017.08.0006.

QGIS Development Team. 2018. QGIS Geographic Information System. Open Source Geospatial Found. Proj. doi: 10.1038/nrg1769.

R Core Team. 2019. R: A Language and Environment for Statistical Computing. https://www.r-project.org/.

Roussel, J.-R., and D. Auty. 2019. lidR: Airborne LiDAR Data Manipulation and Visualization for Forestry Applications. https://cran.r-project.org/package=lidR.

Rutkoski, J., J. Poland, S. Mondal, E. Autrique, L. González Párez, et al. 2016. Canopy Temperature and Vegetation Indices from High-Throughput Phenotyping Improve Accuracy of Pedigree and Genomic Selection for Grain Yield in Wheat. G3 Genes|Genomes|Genetics 6(September): 1–36. doi: 10.1534/g3.116.032888.

Shi, Y., S.C. Murray, W.L. Rooney, J. Valasek, J. Olsenholler, et al. 2016a. Corn and sorghum phenotyping using a fixed-wing UAV-based remote sensing system. In: Valasek, J. and Thomasson, J.A., editors. p. 98660E

Shi, Y., J.A. Thomasson, S.C. Murray, N.A. Pugh, W.L. Rooney, et al. 2016b. Unmanned Aerial Vehicles for High-Throughput Phenotyping and Agronomic Research. PLoS One. doi: 10.1371/journal.pone.0159781.

Souza, C.R. de, A. Gaidon, Y. Cabon, and A.M. Lopez. 2017. Procedural Generation of Videos to Train Deep Action Recognition Networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE. p. 2594–2604

Stanton, C., M.J. Starek, N. Elliott, M. Brewer, M.M. Maeda, et al. 2017. Unmanned aircraft system-derived crop height and normalized difference vegetation index metrics for sorghum yield and aphid stress assessment. J. Appl. Remote Sens. 11(2): 026035. doi: 10.1117/1.JRS.11.026035.

Tattaris, M., M.P. Reynolds, and S.C. Chapman. 2016. A Direct Comparison of Remote Sensing Approaches for High-Throughput Phenotyping in Plant Breeding. Front. Plant Sci. 7(August): 1131. doi: 10.3389/fpls.2016.01131.

Tulsiani, S., A.A. Efros, and J. Malik. 2018. Multi-view Consistency as Supervisory Signal for Learning Shape and Pose Prediction. Computer Vision and Pattern Regognition (CVPR)

Varela, S., P.R. Dhodda, W.H. Hsu, P.V.V. Prasad, Y. Assefa, et al. 2018. Early-season stand count determination in Corn via integration of imagery from unmanned aerial systems (UAS) and supervised learning techniques. Remote Sens. 10(2). doi: 10.3390/rs10020343.

Wang, J., P. Badenhorst, A. Phelan, L. Pembleton, F. Shi, et al. 2019. Using Sensors and Unmanned Aircraft Systems for High-Throughput Phenotyping of Biomass in Perennial Ryegrass Breeding Trials. Front. Plant Sci. 10(October): 1–9. doi: 10.3389/fpls.2019.01381.

White, J.W., P. Andrade-Sanchez, M.A. Gore, K.F. Bronson, T.A. Coffelt, et al. 2012. Field-based phenomics for plant genetics research. F. Crop. Res. 133: 101–112. doi: 10.1016/j.fcr.2012.04.003.

Zaman-Allah, M., O. Vergara, J.L. Araus, A. Tarekegne, C. Magorokosho, et al. 2015. Unmanned aerial platform-based multi-spectral imaging for field phenotyping of maize. Plant Methods 11(1): 35. doi: 10.1186/s13007-015-0078-2.

TABLES

Table 1. Summary of the ANOVA with degrees of freedom (Df) and mean squares for the six analyses utilized for simulation scenario comparison. Values were obtained from 54 simulation scenarios composed by the combination of sample seed (1, 2, and 3) treatment variance (σ_g^2; 0.005 and 0.05), treatment repeatability (R^2; 0.3, 0.6, and 0.9), and flight height (h; 20, 40, and 60). Besides the factors mentioned above, the effect of percentile and resultant interactions were also tested.

Source	Df	Analysis[&]											
		1		2		3		4		5		6	
Seed	2	31.63		1.03	***	0.20	***	0.02		0.00		0.00	
$\boldsymbol{\sigma_g^2}$	1	216.00	***	146.12	***	22.37	***	0.02		0.09	**	1.97	***
$\boldsymbol{R^2}$	2	91.63	**	5.72	***	0.98	***	0.02		0.02		0.02	
$\boldsymbol{h}$	2	29.80		0.16	*	0.03		0.44	***	0.11	***	0.01	
$\boldsymbol{\sigma_g^2{:}R^2}$	2	-		-		0.63	***	-		-		-	
$\boldsymbol{\sigma_g^2{:}h}$	2	-		-		0.00		-		-		-	
$\boldsymbol{R^2{:}h}$	4	-		-		0.03		-		-		-	
$\boldsymbol{p}$	6	-		2.15	***	-		-		-		-	
$\boldsymbol{p{:}\sigma_g^2}$	6	-		0.23	***	-		-		-		-	
$\boldsymbol{p{:}R^2}$	12	-		0.05		-		-		-		-	
$\boldsymbol{p{:}h}$	12	-		0.01		-		-		-		-	
Residuals[#]	-	14.74		0.05		0.01		0.01		0.01		0.01	

[&]The analysis differed on the dependent variable of the linear model which were composed of: 1) the percentiles that maximized the PH_{sim} and PH_{htp} correlation; 2) the correlations between PH_{sim} and PH_{htp} using different percentiles (p = 80, 85, 90, 95, 100, optimal or optimized); 3) the correlations between PH_{sim} and PH_{htp} obtained with the 90th percentile; 4) the values of intercepts obtained from the regression of PH_{sim} on PH_{htp}; 5) the values of the regression coefficient obtained from the regression of PH_{sim} on PH_{htp}; and 6) The Jaccard's index values from the selection based on PH_{sim} and PH_{htp};
Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 ' ' 1;
[#]The number of degrees of freedom of residuals was 46 for analysis 1, 4, 5, and 6, 334 for analysis 2, and 38 for analysis 3.

Table 2. Tukey test for the significant factors identified in the ANOVA. Factors were flight altitude (h; 20, 40, and 60), treatment variance (σ_g^2; 0.005 and 0.05), and treatment repeatability (R^2; 0.3, 0.6, and 0.9).

Source	Level	Analysis				
		1	3	4	5	6
h	20	-	-	-0.26 a	1.01 a	-
	40	-	-	-0.47 b	0.88 b	-
	60	-	-	-0.57 c	0.86 b	-
σ_g^2	0.0005	91.04 a*	0.72 b	-	0.96 a	0.42 b
	0.05	87.04 b	0.97 a	-	0.88 b	0.81 a
R^2	0.3	86.44 b	0.82 c	-	-	-
	0.6	90.11 a	0.85 b	-	-	-
	0.9	90.56 a	0.86 a	-	-	-

*groups with different letters are statistically different at 0.05 probability;
&The analysis differed on the dependent variable of the linear model which were composed of: 1) the percentiles that maximized the PH_{sim} and PH_{htp} correlation; 3) the correlations between PH_{sim} and PH_{htp} obtained with the 90th percentile; 4) the values of intercepts obtained from the regression of PH_{sim} on PH_{htp}; 5) the values of the regression coefficient obtained from the regression of PH_{sim} on PH_{htp}; 6) The Jaccard's index values from the selection based on PH_{sim} and PH_{htp}.

FIGURES

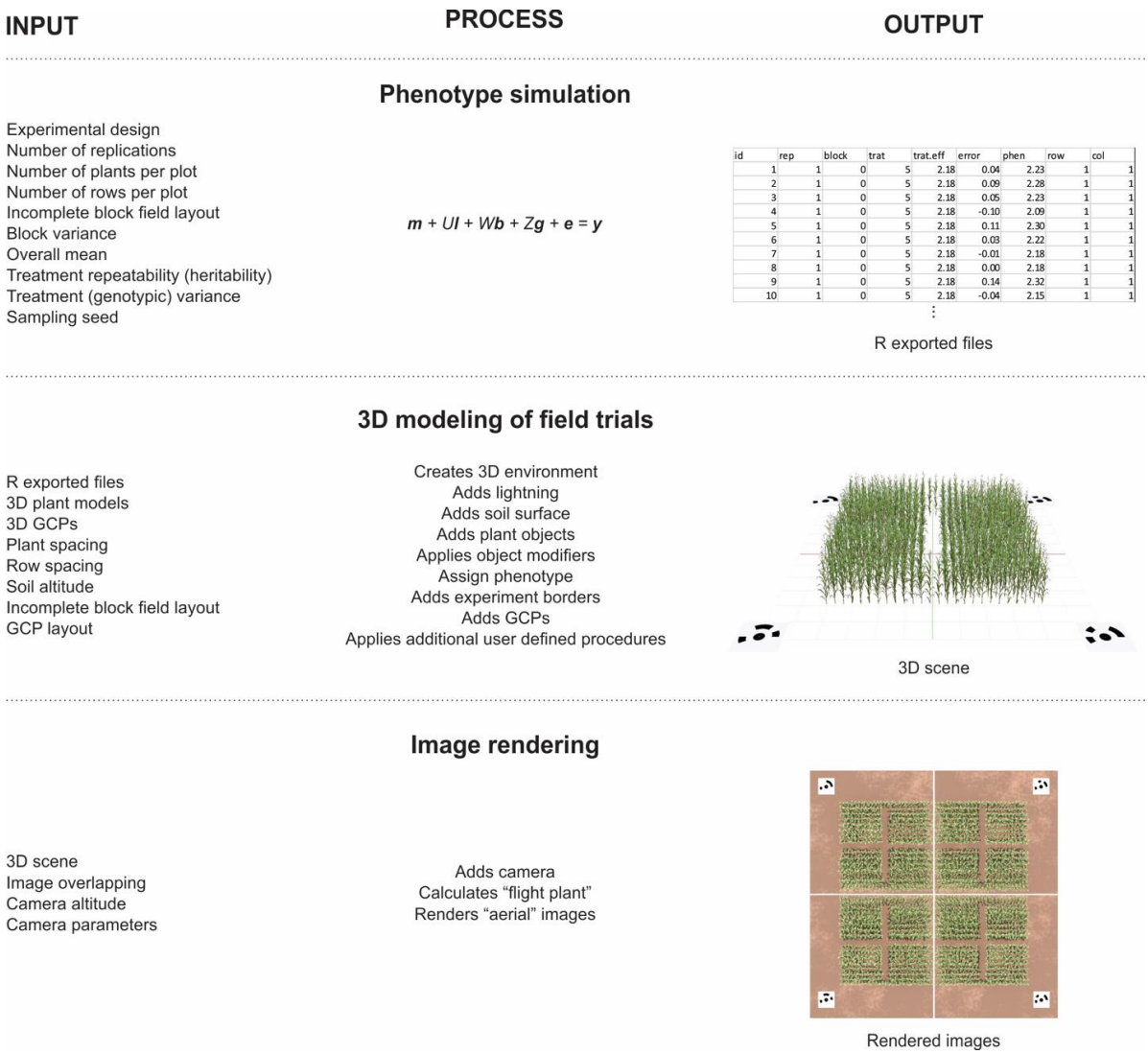

Figure 1. Summary of the inputs, processing steps and outputs of the proposed algorithm. The processing step is composed of phenotype simulation, 3D modeling of the field trial, and image rendering.

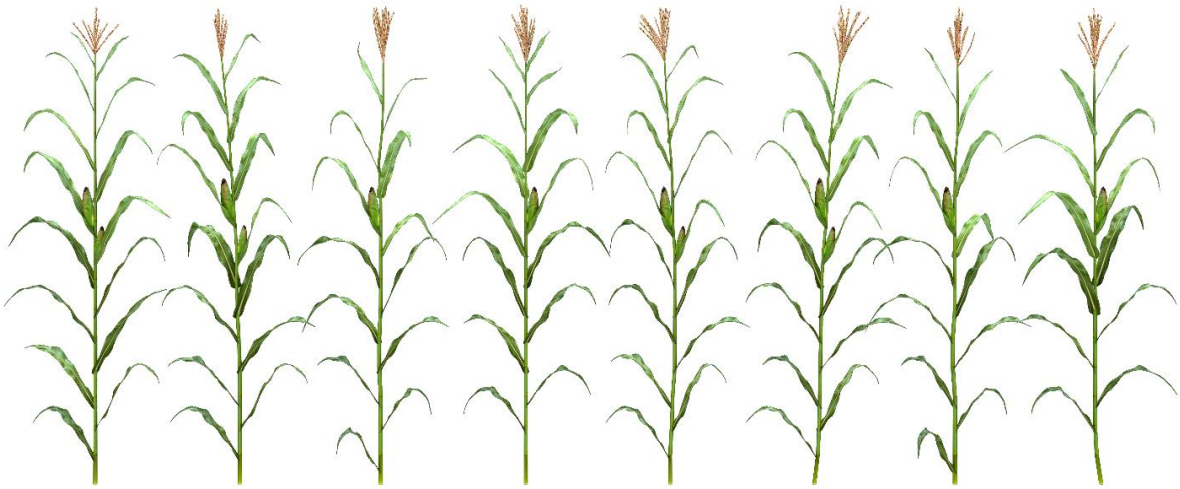

Figure 2. Plant models used in the simulations. The plant is represented by a set of polygons with known spatial coordinates.

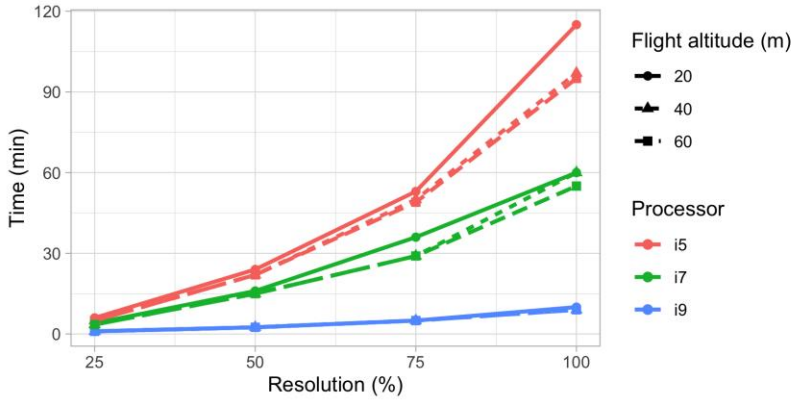

Figure 3. Time (min) required by Blender Cycles' engine to render an experiment image (Figure S3) in four resolutions under three flight altitudes (m) with three different processors (Table S1). Resolutions are presented in percentages (25, 50, 75 and 100%) of 5472×3648 pixels.

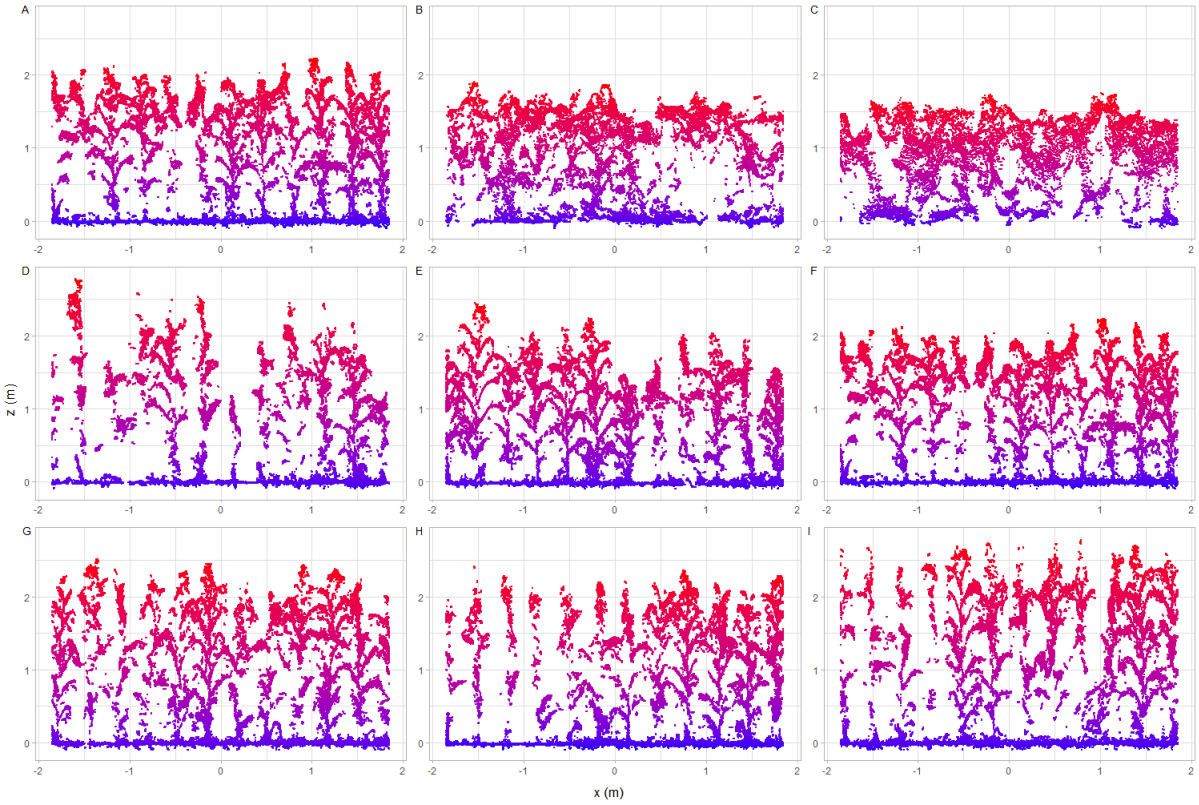

Figure 4. Two-dimensional point cloud representation of reconstructed plots. The x axis represents the extent of the plot and the y axis the height of the points. A, B, and C represent the same experimental unit (σ_g^2 = 0.05, R^2 = 0.9, replication = 1, and seed = 1) rendered at different heights (A:20, B:40, and C:60). D, E, and F are experimental units of the same treatment sampled from the set of high variance population (σ_g^2 = 0.05) considering the same seed (1), replication (1), and flight height (20) with varying repeatabilities (D: 0.3, E: 0.6, and F: 0.9). That is, the difference between these plots is the magnitude of the errors. G, H, and I are the same as D, E, and F, but with for a treatment drawn from the treatments simulated with low treatment variance (σ_g^2 = 0.0005).

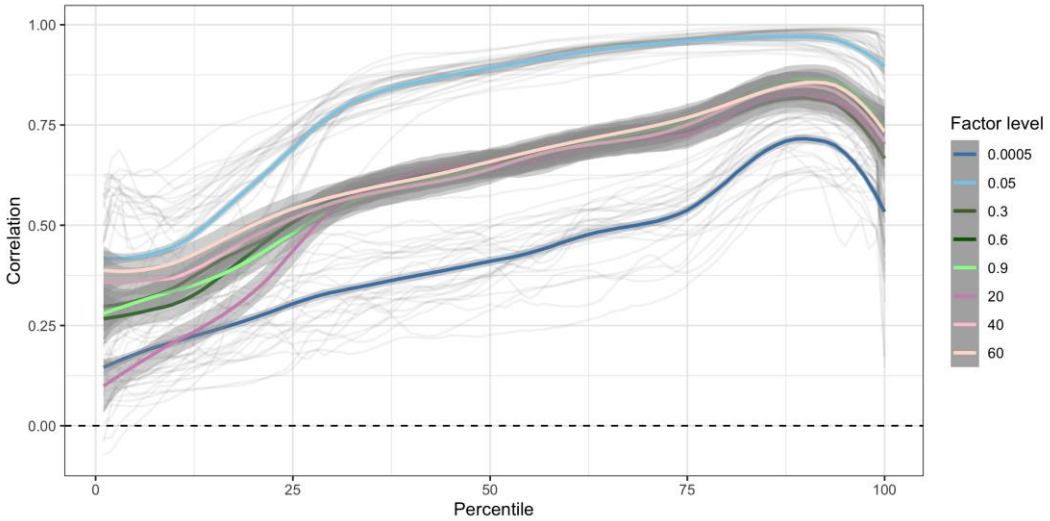

Figure 5. Correlation between simulated and estimated heights for percentiles varying from 1 to 100. Simulations were composed of the combination of different treatment variance levels (0.0005 and 0.05), repeatabilities (0.3, 0.6, and 0.9), and flight heights (20, 40, 60). Each simulation was replicated three times using sampling seeds. Gray lines represent the correlation of each replicated scenario and colored lines represent the local regression for each main level of the simulations.

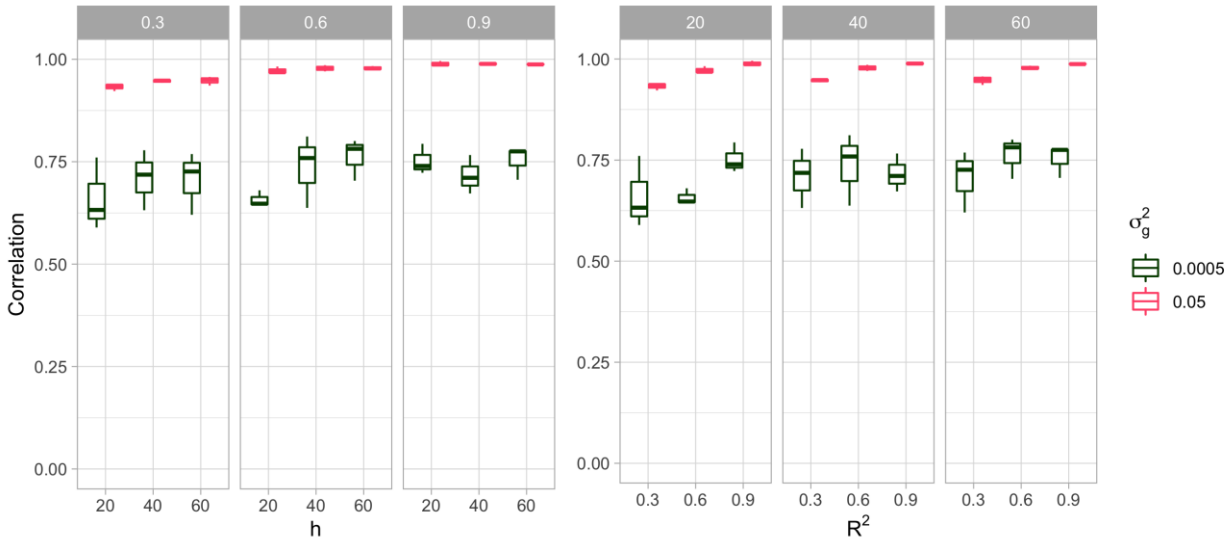

Figure 6. Boxplots of correlations between simulated and HTP estimated heights for 54 simulated scenarios composed by combining treatment variance (σ_g^2), flight altitude (h), treatment repeatability (R^2), and sampling seed. The graph on the left is indexed by repeatability, and on the right is indexed by flight altitude.

SUPPLEMENTARY TABLES

Table S1. Processors tested on the rendering process of Blender's Cycles engine.

Processor	Operational system	RAM size (GB)
Intel Core i5 Dual-Core @2.7 GHz	macOS Catalina	8
Intel Core i7-3770 @3.5 GHz	Windows 10 Pro	32
Intel Core i9-7980XE @2.6 GHz	Windows 10 Pro	128

Table S2. Procedures and parameters utilized in Metashape for reconstructing the computed generated experiment.

Procedure		Automation	Parameter	Value
1	Add photos	Manual	-	-
2	Import photo coordinates	Manual	-	-
3	Camera calibration	Batch	*f*	9949.09105
4	Estimate camera quality	Batch	Remove if less than	0.70
5	Align cameras	Batch	Accuracy	Highest
			Generic preselection	Yes
			Reference preselection	Yes
			Key point limit	40,000
			Tie point limit	40,000
			Adap. cam. model fitting	No
6	Build mesh (batch)	Batch	Source data	Sparse cloud
			Surface type	Arbitrary 3D
			Face count	High
7	Import and place GCP	Automatic	Tolerance	60
8	Optimize camera alignment	Batch	General	*f, cx, cy, k1-k3, p1, p2*
9	Set bounding box	Manual	-	-
10	Dense cloud	Batch	Quality	High
			Depth filtering	Moderate
11	Clean dense cloud (if required)	Manual	-	-
12	Build digital elevation model	Batch	Source data	Dense cloud
			Interpolation	Enabled
13	Build orthomosaics	Batch	Blending mode	Mosaic
			Surface	DEM
			Enable hole filling	Yes

Table S3. Survey statistics for the 54 reconstructed scenarios based on combinations of seed (1, 2, and 3), treatment variation levels (σ_g^2; 0.0005 and 0.05), repeatabilities (R^2; 0.3, 0.6, and 0.9), and flight altitudes (20, 40, 60).

Seed	R^2	σ_g^2	Altitude (m)	Est. flying altitude (m)	Ground res. (mm pix^{-1})	Reprojection error (pix)	GCP error (mm; RMSE)		
							x	y	z
1	0.3	0.0005	20	25.3	1.99	1.08	48.28	50.36	5.33
			40	42.3	4.01	0.43	2.44	5.73	0.68
			60	57.7	6.01	0.80	1.35	6.56	0.31
		0.05	20	23.8	2.00	1.10	65.02	43.28	8.15
			40	39.5	4.01	0.47	1.43	6.49	0.69
			60	63.5	6.01	0.80	1.35	6.58	0.20
	0.6	0.0005	20	24.5	2.00	1.05	14.08	16.70	40.73
			40	41.7	4.01	0.47	1.59	6.37	0.99
			60	59.2	6.01	0.42	1.34	6.66	0.21
		0.05	20	21.2	2.00	1.02	31.12	46.40	54.97
			40	42.1	4.01	0.47	1.33	6.07	0.56
			60	56.3	6.01	0.80	1.34	6.59	0.42
	0.9	0.0005	20	28.6	2.00	1.05	3.07	6.86	0.34
			40	41.6	4.01	0.88	1.48	6.43	0.71
			60	57.5	6.01	0.80	1.36	6.55	0.46
		0.05	20	23.8	2.00	1.05	26.49	38.00	0.64
			40	42.9	4.01	0.47	2.46	5.73	1.01
			60	63.0	6.01	0.42	1.43	6.58	0.26
2	0.3	0.0005	20	25.0	1.99	1.12	37.85	35.37	54.65
			40	42.1	4.01	0.47	1.35	6.61	0.54
			60	59.0	6.01	0.42	1.32	6.68	0.23
		0.05	20	25.2	1.99	1.07	54.50	48.20	3.84
			40	41.8	4.01	0.47	1.36	6.60	0.62
			60	59.2	6.01	0.42	1.34	6.65	0.12
	0.6	0.0005	20	24.4	2.00	1.03	3.05	6.58	2.19
			40	40.6	4.01	0.47	2.37	5.71	1.17
			60	55.2	6.01	0.43	1.33	6.61	0.08
		0.05	20	24.5	2.00	1.09	33.25	43.69	54.24
			40	45.5	4.01	0.48	1.38	6.49	0.81
			60	64.2	6.01	0.80	1.35	6.65	0.15
	0.9	0.0005	20	24.3	2.00	1.07	42.95	31.99	3.73
			40	37.8	4.01	0.86	1.36	6.45	0.56
			60	59.5	6.01	0.42	1.35	6.67	0.14
		0.05	20	23.8	2.00	1.09	45.22	39.83	3.93
			40	42.2	4.01	0.47	2.47	5.74	0.67
			60	59.6	6.01	0.42	1.33	6.65	0.23
3	0.3	0.0005	20	21.6	2.00	1.05	21.03	19.65	64.78
			40	42.1	4.01	0.47	1.68	6.14	2.04
			60	62.5	6.01	0.42	1.38	6.78	1.48
		0.05	20	21.2	2.00	1.02	47.39	66.74	52.16
			40	41.5	4.01	0.47	2.55	5.77	0.72
			60	62.2	6.01	0.42	1.32	6.72	0.30
	0.6	0.0005	20	21.3	2.00	1.03	7.41	9.57	62.89
			40	42.0	4.01	0.47	1.85	7.01	0.94
			60	61.9	6.01	0.42	2.62	5.81	0.18
		0.05	20	21.3	2.00	1.02	18.35	23.07	76.09
			40	40.5	4.01	0.86	1.29	6.61	0.49
			60	62.3	6.01	0.42	1.32	6.70	1.57
	0.9	0.0005	20	21.2	1.99	1.03	38.93	32.60	40.45
			40	42.7	4.01	0.47	1.30	6.51	1.37
			60	63.2	6.01	0.80	1.31	6.69	0.41
		0.05	20	21.5	1.99	1.04	1.45	6.54	49.34
			40	41.8	4.01	0.47	3.93	5.29	3.82
			60	58.9	6.01	0.81	1.33	6.66	0.24

Table S4. Tukey test for the effect of percentile (p = 80, 85, 90, 95, 100, optimal or optimized) on the correlations between PH_{sim} and PH_{htp}.

Level	Mean (group)*	
optimal	0.85	a
90	0.84	a
85	0.83	ab
95	0.82	ab
80	0.79	bc
optimized	0.78	c
100	0.67	d

*groups with different letters are statistically different at 0.05 probability.

Table S5. Summary of ANOVAs by treatment variation (σ_g^2; 0.0005 and 0.05) testing the effect of seed, treatment repeatability (R^2) and flight altitude (h) for the correlations between simulated and estimated plant heights using the 90th percentile.

σ_g^2	Source	Df	Mean Squares	F value	
0.0005	Seed	2	0.10	12.14	**
	R^2	2	0.02	2.43	
	h	2	0.02	2.85	
	Residuals	20	0.01		
0.05	Seed	2	0.11	6.94	**
	R^2	2	1.59	97.16	***
	h	2	0.01	0.49	
	Residuals	20	0.02	-	

Df: degrees of freedom;
Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 ' ' 1.

SUPPLEMENTARY FIGURES

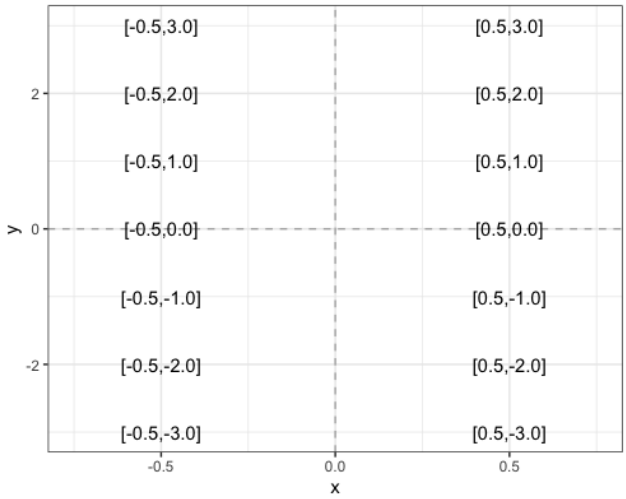

Figure S1. Experiment layout for the simulated experiment. The coordinates $[x, y]$ of block centers were [-0.5,3.0], [0.5,3.0], [-0.5,2.0], [0.5,2.0], [-0.5,1.0], [0.5,1.0], [-0.5,0.0], [0.5,0.0], [-0.5,-1.0], [0.5,-1.0], [-0.5,-2.0], [0.5,-2.0], [-0.5,-3.0] and [0.5,-3.0].

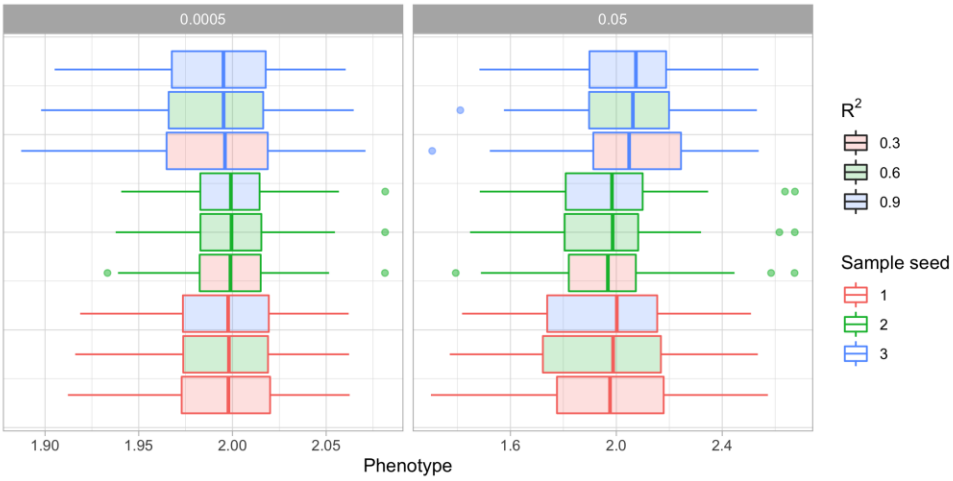

Figure S2. Boxplot of simulated phenotypes (plant height) indexed by treatment variation (σ_g^2; 0.0005 and 0.05), treatment repeatability (R^2; 0.3, 0.6, and 0.9) and sample seed (1, 2, and 3). The scales are different for each treatment variation.

Figure S3. Images utilized to test rendering time of Blender's Cycles engine at A) 20 m; B) 40 m; and C) 60 m.

Figure S4. Exemplification of a simulated scenario (h = 40, σ_g^2 = 0.0005; R^2 = 0.9; and seed = 3) showing the outputs obtained from the rendered images. A) Example of *in silico* images rendered in Blender's Cycles engine; B) Top-view of the reconstructed point cloud obtained from structure from motion; C) reconstructed DSM; D) reconstructed orthomosaic; E) ground-truth orthophoto obtained in Blender (presented for reference).

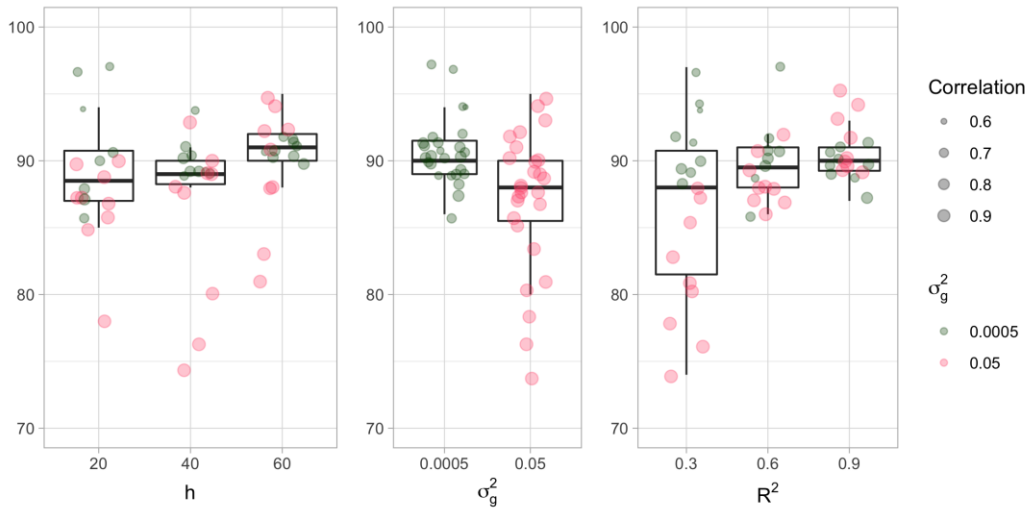

Figure S5. Bubble graph with the percentile that generates maximum correlation between ground truth and HTP estimated heights as a function of treatment variation levels (σ_g^2; 0.0005 and 0.05), repeatabilities (R^2; 0.3, 0.6, and 0.9), and flight altitudes (h; 20, 40, 60). The size of the points is a function of the magnitude of the correlations. The points are colored by treatment variation.

4. GENERAL CONCLUSIONS

We reported on two studies regarding the optimization and validation of UAS-based HTP. Our findings indicated the possibility of optimizing HTP based on the identification of the most informative flight dates and secondary traits. The results suggest that using a single flight date (119 DAS) and secondary trait (NDVI or MSR) can maximize prediction accuracy for both grain yield and plant health (anthracnose resistance) in sorghum hybrids.

Furthermore, we proposed a novel way to validate and design UAS HTP pipelines based on computer-generated simulations. Our results showed that synthetic experiments can be adequately reconstructed from rendered images using structure-from-motion algorithms. Using this tool, we inferred that treatment variance, experimental repeatability and the chosen percentile of points from dense clouds affect the accuracy of HTP.