






# Temporal field phenomics of transgenic maize events subjected to drought stress: Cross-validation scenarios and machine learning models

Helcio Duarte Pereira<sup>1,2</sup>  | Juliana Vieira Almeida Nonato<sup>1,2</sup>  |  
 Rafaela Caroline Rangni Moltocar Duarte<sup>1,2,3</sup>  | Isabel Rodrigues Gerhardt<sup>1,2,4</sup>  |  
 Ricardo Augusto Dante<sup>1,2,4</sup>  | Paulo Arruda<sup>1,2,5</sup>  |  
 Juliana Erika de Carvalho Teixeira Yassitepe<sup>1,2,4</sup> 

<sup>1</sup>Genomics for Climate Change Research Center (GCCRC), Universidade Estadual de Campinas, Campinas, São Paulo, Brazil

<sup>2</sup>Centro de Biologia Molecular e Engenharia Genética (CBMEG), Universidade Estadual de Campinas, Campinas, São Paulo, Brazil

<sup>3</sup>Embrapa Meio Ambiente, Jaguariúna, São Paulo, Brazil

<sup>4</sup>Embrapa Agricultura Digital, Campinas, São Paulo, Brazil

<sup>5</sup>Departamento de Genética, Evolução, Microbiologia e Imunologia, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, São Paulo, Brazil

## Correspondence

Juliana Erika de Carvalho Teixeira Yassitepe, Genomics for Climate Change Research Center (GCCRC), Universidade Estadual de Campinas, Campinas, São Paulo, Brazil.

Email: [juliana.yassitepe@embrapa.br](mailto:juliana.yassitepe@embrapa.br)

Assigned to Associate Editor Sara Tirado Tolosa.

## Funding information

Fundação de Amparo à Pesquisa do Estado de São Paulo, Grant/Award Numbers: 2016/23218-0, 2022/04930-1, 2023/11640-2

## Abstract

Global climate change has driven breeding programs to develop abiotic stress-resilient plant varieties. Traditionally, assessing drought resilience involves labor-intensive and time-consuming processes. This study used an unmanned aerial system (UAS) to predict key phenotyping traits in maize (*Zea mays* L.) and monitor plant response to drought during the crop cycle. We grew transgenic maize hybrids in two trials, one irrigated and another subjected to drought stress, and used a drone equipped with red–green–blue (RGB) and multispectral sensors to capture images of the plots over time. Machine learning models and various prediction scenarios revealed significant correlations between vegetation indices over time. Interestingly, the RGB sensor outperformed the multispectral sensor in trait prediction. Prediction accuracy across scenarios with untested genotypes and environments ranged from 0.40 to 0.70 for grain yield, 0.43 to 0.69 for days to anthesis, 0.51 to 0.67 for days to silking, and 0.35 to 0.57 for plant height. Ridge and random forest models consistently delivered the most accurate predictions across traits and environments. The vegetation indices normalized green–red difference index, VARI, and RCC also effectively predicted and captured the plant response to drought. This study highlights the value of UAS phenotyping as a practical tool for assessing abiotic stress due to its straightforward implementation.

## Plain Language Summary

Climate change has become a new threat to crop production, and drought is among the greatest concerns for sustaining food production. Plant breeding plays a special role in developing new and more resilient genotypes. Since evaluations under stressful

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2025 The Author(s). *The Plant Phenome Journal* published by Wiley Periodicals LLC on behalf of American Society of Agronomy and Crop Science Society of America.

conditions require extensive assessments, a new tool to aid breeders was investigated in this study. A UAS with integrated RGB and multispectral sensors was employed to capture images at all phenological stages of the maize crop cycle in two experiments: one irrigated and the other under drought stress. These images were used to monitor the plant response to drought and predict important traits. The prediction accuracies achieved by machine learning models using RGB data showed promising applications. Some vegetation indices reflected how plants changed under drought and correlated highly with the assessed traits. Thus, digital phenotyping can be a new tool to obtain reliable phenotypes more quickly and affordably.

## 1 | INTRODUCTION

Maize (*Zea mays* L.) is grown worldwide and is crucial in ensuring global food security (Erenstein et al., 2022; Zheng et al., 2018). As the world's population continues to grow, the demand for food and resources has increased, leading to a steady rise in global maize production (FAOSTAT, 2022). However, global climate change is posing new challenges to crop production, with drought, heat, and their interactions being the most significant threats to agricultural systems (Lobell et al., 2011, 2013; Wang et al., 2021). Over time, the frequency and severity of water stress are expected to increase (Yuan et al., 2023).

Plant responses to drought are characterized by complex and interrelated mechanisms that are not yet fully understood (Li et al., 2023). Factors such as phenological stage, intensity, duration, and interaction with other stresses contribute to this complexity. Developing drought-tolerant cultivars can be expensive and time-consuming (McMillen et al., 2022; Messina et al., 2021). The development and validation of new tools to assist breeders in creating stress-tolerant cultivars are in high demand, especially for annual grain crops with shorter cycles and high replacement rates by new cultivars.

The response to drought involves physiological and morphological adaptations throughout the plant life cycle (Yan et al., 2023). To fully understand this phenomenon, multiple phenological aspects must be assessed over time, requiring labor expertise that is not widely affordable to many research groups. Tools that allow for repeated, fast, and accessible recording over time, such as unmanned aerial systems (UASs) phenotyping, can contribute to a more comprehensive understanding of plant responses to drought stress (Araus & Cairns, 2014). UAS has been applied in a low-cost, high-throughput manner for phenomic selection in wheat (*Triticum aestivum* L.) and poplar (*Populus nigra* L.), providing numerous variables that can be used as regressors or to estimate kinship in the statistical models used in genomic selection (Rincent et al., 2018). With advancements in sensors, unmanned aerial vehicles (UAVs), geographic information systems, and image

processing software, phenomic prediction has increasingly been applied in field trials (Aguate et al., 2017; Anderson et al., 2019).

Understanding how new plant genotypes behave in different environments is challenging for plant breeding programs. The interaction effects between genotype and environment are usually significant for most crops. Due to limited resources and seeds, evaluating all target environments is impossible, which is why prediction is so important. Researchers have looked into the potential of phenomic prediction to be used across different environments or for unassessed genotypes since its first applications (Lane et al., 2020; Montesinos-López et al., 2017; Wijewardane et al., 2023). Designing appropriate cross-validation (CV), namely, predicting incomplete field trials, new genotypes, or new target environments that have not been observed in the field, can allow the breeding program to advance while still using the same amount of plots (Jarquin et al., 2017; Pandey et al., 2020). The CV scheme, aligned with a powerful prediction tool, will borrow genetic relationships or phenotypic correlations among environments and can aid in increasing the accuracy of selection in early screenings (Jarquin et al., 2020).

Compared to genomic selection, a well-established tool in breeding, phenomic prediction has generally demonstrated superior performance across diverse environments, largely due to its ability to capture nonadditive effects (Robert et al., 2022; Winn et al., 2023). It is also less affected by population structure and kinship (Weiß et al., 2022) and the size of the training population (Zhu et al., 2022). On the other hand, genomic selection relies on fixed information from molecular markers that are not influenced by environmental conditions. Different from genomic selection models that are largely based on a mixed model framework (RRBLUP, GBLUP) or Bayesian inference (Bayes-A, Bayes-B, Bayes-least absolute shrinkage operator [LASSO], etc.), the phenomic prediction models are mainly based on machine learning techniques, such as penalized regressions (elastic net, ridge regression [Ridge], LASSO regression), decision trees (random forest [RF]), and variable selection approaches (partial least

squares [PLS]) (Chafai et al., 2023). Machine learning models outperformed their genomic counterparts in predicting grain yield (GY) and grain protein content using spectral information in a wheat breeding program (Sandhu et al., 2021).

Using phenomic prediction for quantitative traits in initial-stage field trials is an emerging trend in plant breeding programs. In the early generations, the limited number of seeds forced breeders to use small plots and fewer replications/environments, challenging traits like GY to be assessed. However, small-field wheat trials have observed good correlations between phenomic predictions and GY (Krause et al., 2020). Similarly, developing biotech traits encounters obstacles, including a scarcity of genetically engineered seeds, high trial costs, and the necessity to rely on highly heritable secondary traits related to GY for effective field evaluation. In light of these challenges, our study aims to (1) investigate the potential of phenomic prediction for early screenings, including assessments under drought conditions, and (2) identify the most effective vegetation indices for predicting primary traits and evaluating water stress. The study also examined the impact of sensors, prediction scenarios, and machine-learning regressions on these predictions.

## 2 | MATERIALS AND METHODS

### 2.1 | Genotypes and experimental design

Two trials involving 21 F<sub>1</sub> hybrids, including 18 transgenic events overexpressing drought-tolerance candidate genes and three non-transgenic genotypes (Table S1), were planted in Campinas, SP, Brazil (22°54'23" S 47°3'42" W) during the dry season (April–September) of 2023. These transgenic events consisted of different genes and genetic backgrounds (temperate, tropical, and adapted) selected based on their demonstrated drought tolerance in controlled environments, including growth chambers and greenhouses. The genotypes were assessed using a randomized complete blocks design with three replications. The experimental plots were set up in single rows with a length of 5 m and spaced 0.45 m apart. The soil was prepared through conventional tillage, and all crop management practices were consistent between the two trials, except for the water supply. The trial area had been uncultivated for the previous 3 years and had a uniform and flat surface.

An inline drip irrigation system was used to water the plants until they reached the V6 stage. After that stage, only the trial with irrigation continued to receive water until the R5 stage. Tensiometers were installed at 20- and 40-cm depths to manage the irrigation system. Each experimental area consisted of 20 rows of a commercial single-cross hybrid planted around as a border.

### Core Ideas

- Temporal imaging captures most vegetation indices variation throughout the maize plant cycle.
- Vegetation indices show low to moderate correlation between and within each imaging time.
- Red–green–blue sensors provide more accurate predictions than the multispectral sensor.
- Ridge and random forest models yield consistent predictions across traits and scenarios.

### 2.2 | Plant imaging and trait extraction

We used a DJI Mavic 3 M UAV (drone) harboring two sensors—red–green–blue (RGB) and multispectral. The drone flew at a height of 12 m and captured RGB images (5280 × 3956 pixels) using a standard integrated camera and multispectral images (2592 × 1944 pixels) with the following bands: green (560 ± 16 nm), red (650 ± 16 nm), red-edge (730 ± 16 nm), and near-infrared (860 ± 26 nm). The frontal and side overlap between images were 90% and 80%, respectively. Each flight lasted around 10 min, and we recorded 290 images during each flight. The flights were conducted between 11 a.m. and 1 p.m. to ensure perpendicular solar incidence, clear skies, and minimal wind conditions whenever possible.

The raw images were processed using the software Open Drone Map (WebODM interface) to generate the orthomosaic. In the main user interface, the RGB images were processed with the options *High Resolution* and *No Resize Images*, and the multispectral images with *multispectral* and *radiometric calibration: camera+sun*. With these settings and the flight parameters, the average ground sampling distance was approximately 0.5 cm for RGB and 0.7 cm for multispectral orthomosaics.

From all flights, we selected the ones without blurred or cloudy areas. Eighteen flights with the RGB sensor and 13 with the multispectral sensor, covering the entire crop cycle, were chosen for further analysis (Figure 1).

The polygons for each plot were set using the QGIS software (QGIS Development Team, 2024) to create ESRI shape files (.shp file extensions) for each trial on each flight date. A fixed percentage buffer of 80% was applied using the *Buffer by percentage* plugin to minimize the influence and overlap of adjacent plots. Once a shape file was created, it was adjusted manually to ensure the correct position over the plots for another orthomosaic and then saved. These shape files of the plots were later utilized to extract the vegetation indices in each trial and flight date.

The orthomosaic and shape file archives were imported into R to obtain the plot-based vegetation indices. The first step



**FIGURE 1** Flight dates as days after planting (DAP) selected for each sensor and the corresponding phase of the crop cycle in which they were located. RGB, red–green–blue.

in the workflow was to clip all the plots from the orthomosaic using an R script with the *terra* package (Hijmans, 2024), which helped save memory and processing time. A color mask was then applied to separate the plants from the soil using the *field mask* function from the *FIELDimageR* R package (Matias et al., 2020), using the overall hue index and the normalized difference vegetation index (threshold 0.6) for RGB and multispectral orthomosaics, respectively.

A total of 630 ( $35 \times 18$ ) and 702 ( $54 \times 13$ ) temporal phenotypes were calculated, including 35 RGB and 54 multispectral vegetation indices (also considering the single bands of each sensor) for each flight date from imaging phenotyping. All calculations of the vegetation indices were performed using an R script with the *terra* package. The study's vegetation indices and their expressions and references can be found in Tables S2 and S3 for RGB and multispectral sensors, respectively. Each plot-based vegetation index used the trimmed mean, discarding 10% of the pixels with extreme values to remove any noise pixels that could remain from soil segregation.

### 2.3 | Statistical analysis of ground and imaging phenotyping

We studied 10 traits related to plant growth and production, including anthesis silking interval (ASI), days to anthesis (DTA), days to silking (DTS), ear diameter (ED), ear height (EH), ear index (EI), ear length (EL), GY, 100-grain weight (HGW), and plant height (PH). The flowering time was recorded when 50% of the plants were shedding pollen or when the first silk emerged. ASI was calculated as the difference between these two dates. PH and EH were measured as the distance from the soil to the flag leaf and the insertion of the first ear, respectively. ED was measured as the diameter at the middle of the ear, while EL was measured as the distance between the base and the tip of the ear. EI was determined as the ratio between the total number of ears and the number of plants in the plot. HGW was calculated as the mean weight of 100 kernels per plot, and GY was determined by weigh-

ing all the plot's threshed ears and adjusting the moisture to 14%.

The data from temporal phenotypic imaging were analyzed using a nested design to obtain temporal best linear unbiased predictions for each genotype. A fully random model was fitted using the *sommer* R package (Covarrubias-Pazaran, 2016), with variance components estimated by restricted maximum likelihood and genetic effects estimated by best linear unbiased prediction (BLUP). Each temporal vegetation index was modeled for both sensors as follows:

$$Y_{ijk} = \mu + \varphi_i + \beta_{i(j)} + \theta_{i(k)} + \varepsilon_{ijk} \quad (1)$$

where  $Y_{ijk}$  is the vegetation index value for the genotype  $j$  in the replication  $k$  at each flight  $i$ , given as days after planting (DAP);  $\mu$  is the overall mean;  $\varphi_i$  is the random effect of  $i$ th flight time with  $\varphi_i \stackrel{iid}{\sim} N(0, \sigma_{\varphi_i}^2)$ ,  $i \in \{26, 35, 40, 49, 55, 60, 67, 74, 91, 95, 99, 104, 109, 113, 119, 126, 131, 139\}$  for RGB sensor, and  $i \in \{26, 35, 40, 49, 55, 74, 76, 95, 99, 104, 110, 113, 123\}$  for multispectral sensor;  $\beta_{i(j)}$  is the random effect of  $j$ th genotype within the  $i$ th flight time with  $\beta_{i(j)} \stackrel{iid}{\sim} N(0, \sigma_{\beta_{i(j)}}^2)$ ,  $j \in \{1, 2, 3 \dots 21\}$ ;  $\theta_{i(k)}$  is the random effect of  $k$ th replication within the  $i$ th flight time with  $\theta_{i(k)} \stackrel{iid}{\sim} N(0, \sigma_{\theta_{i(k)}}^2)$ ,  $k \in \{1, 2, 3\}$ ; and  $\varepsilon_{ijk}$  is the residual pooled error with  $\varepsilon_{ijk} \stackrel{iid}{\sim} N(0, \sigma_{\varepsilon_{ijk}}^2)$ . Temporal repeatability of the genotypic effect was calculated by Equation (2) for each vegetation index.

$$\text{Temporal repeatability} = \frac{\sigma_{\beta_{i(j)}}^2}{\sigma_{\beta_{i(j)}}^2 + \frac{\sigma_{\varepsilon_{ijk}}^2}{\text{number of replications}}} \quad (2)$$

where  $\sigma_{\beta_{i(j)}}^2$  is the genotypic variance regarding all flights and  $\sigma_{\varepsilon_{ijk}}^2$  is the residual error variance.

The same previous model was used with the ground phenotypic data, without the flight component, as we have single time-measured traits, as follows:

$$Y_{jk} = \mu + \beta_j + \theta_k + \varepsilon_{jk} \quad (3)$$

The traditional heritability was calculated based on the variance components from Equation (3), as follows:

$$\text{Heritability} = \frac{\sigma_{\beta_j}^2}{\sigma_{\beta_j}^2 + \frac{\sigma_{\varepsilon_{jk}}^2}{\text{number of replications}}} \quad (4)$$

where  $\sigma_{\beta_j}^2$  is the genotypic variance and  $\sigma_{\varepsilon_{jk}}^2$  is the residual error variance.

The  $R^2$  (total variance explained by the model) for both the imaging phenotyping and traditional ground phenotyping analyses was calculated using the conditional  $R^2$  approach for mixed models, as presented by Nakagawa and Schielzeth (2013). We studied the relative difference between the mean of each vegetation index in the drought and irrigated trials to identify the most sensitive among them. This difference accounted for the vegetative stage, excluding the possible influence of flowering organs or natural leaf senescence. Figure 2 provides an overview of the field-based imaging phenotyping process.

## 2.4 | Prediction scenarios and machine learning regression

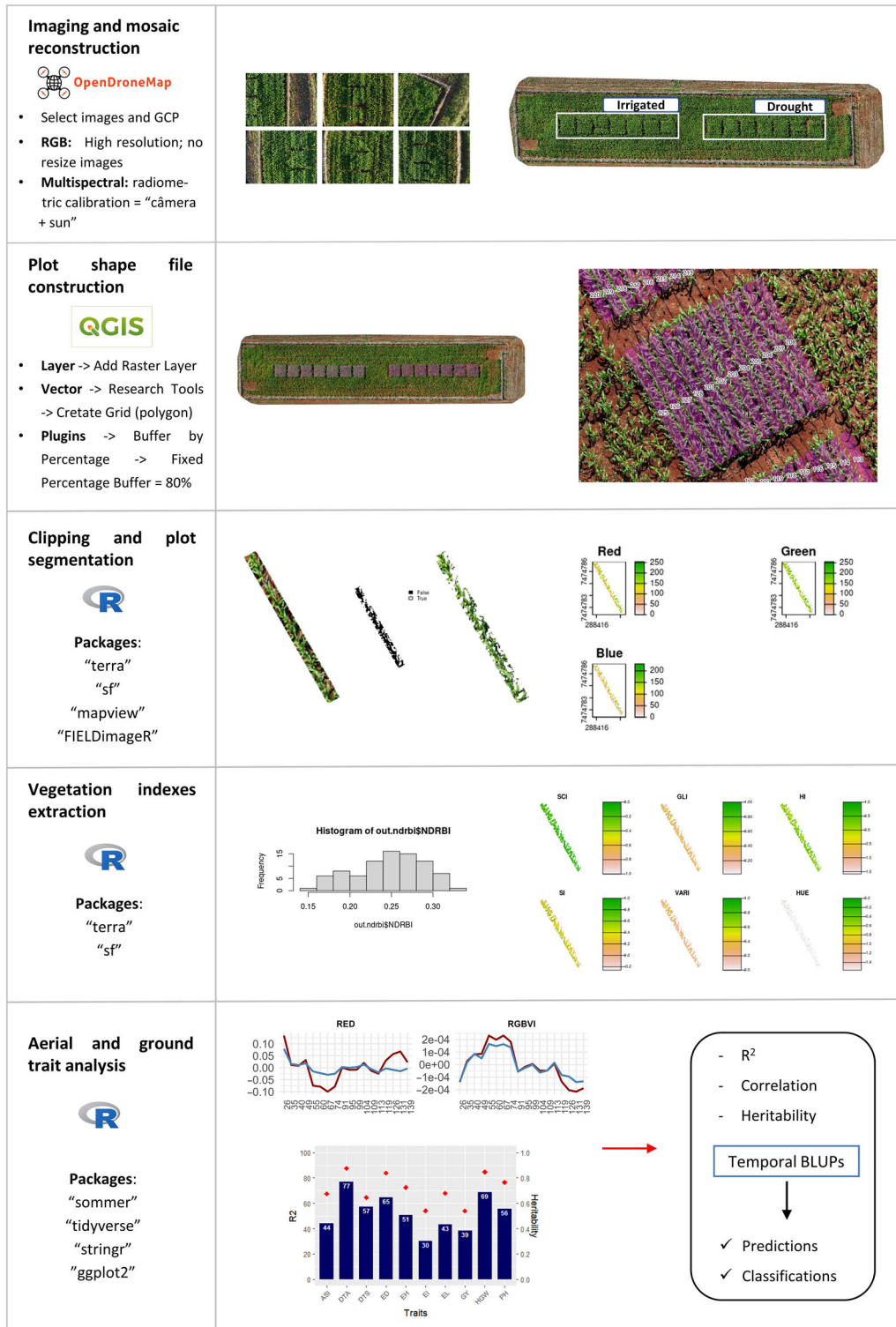
The temporally adjusted breeding values for all RGB and multispectral vegetation indices across the flight dates were used to predict ground phenotyping traits (predicted variables) using machine learning regressions, adjusted in the *caret* R package (Kuhn, 2008). The indices GR and VEG in the RGB data (Table S2) and the indices MNDRE, MRETVI, REGNDVI, REGRVI, RENDVI, and RESR in the multispectral data (Table S3) were excluded from further analyses due to their minimal genotypic variance, identified by the *caret::nearZeroVar* function. Hence, predictions based on RGB and multispectral data used 594 (630 – 2 × 18) and 625 (702 – 6 × 13) imaging phenotyping data, respectively. The genotypes were divided into training and validation datasets of 80% and 20%, respectively, in 500 bootstrap resampling. It is important to note that the same division was employed within each iteration to ensure fair comparisons. Four prediction scenarios mimicking situations that breeders face in plant breeding were examined, encompassing the two prediction populations and the water condition of the trials, similar to that applied in genomic prediction for multi-environments.

The training dataset was the set of genotypes assessed in the field, whose phenotypic records were used to calibrate the models. These genotypes were called tested individuals. The validation dataset was the set of genotypes not assessed

in the field and lacking phenotypic records. These genotypes were called untested individuals. A similar division was also applied to the environments. Thus, we had the environment where the genotypes were assessed, called the tested environment, and the environment where the genotypes were not assessed, called the untested environment. The irrigated trial was used as the tested environment where the genotypes were evaluated, while the drought trial was the untested environment for predicting the genotypes. There were four scenarios for prediction: 1) using the training set in the irrigated trial for tested genotypes in the tested environment (CV1); 2) using the validation set in the irrigated trial for untested genotypes in the tested environment (CV2); 3) using the training set in the drought trial for tested genotypes in the untested environment (CV3), and 4) using the validation set in the drought trial for untested genotypes in the untested environment (CV4). CV1 was used as a benchmark, tested genotypes in a tested environment; CV2 resembles incomplete field trial within a tested environment; CV3 refers to cases where tested genotypes are predicted in new environments; CV4 refers to cases where untested genotypes are predicted in new environments. Figure 3 shows a visual representation of the prediction pipeline.

We explored eight machine learning regression methods commonly used in phenomic prediction research with the *caret::train* function. The *method* argument was set as “glmnet” for Ridge, lasso regression, and elastic net; “knn” for  $k$ -nearest neighbor (KNN); “pls” for partial least squares; “rf” for random forest; “svmLinear” for linear kernel support vector machine (SVM); and “svmRadial” for radial kernel SVM. For tuning parameters, we set *alpha* as 0 for Ridge, 1 for lasso, and a searched value between 0 and 1 for elastic net by 10 equal increments. We defined *lambda* as a searched value between 0.0001 and 1 in 10 equal increments for ridge, lasso, and elastic net. To tune the parameters for PLS and KNN, we set *tuneLength* as 10 for the number of principal components and 100 for the number of  $k$ , respectively. For RF, we defined *nTree* (number of trees) as 1000 and *mtry* (number of variables tested in each split) as a searched value between 1 and 50. For the SVM linear, we defined *tuneGrid* as a search value between 1 and 5, and for the SVM radial, we set *tuneLength* to 10.

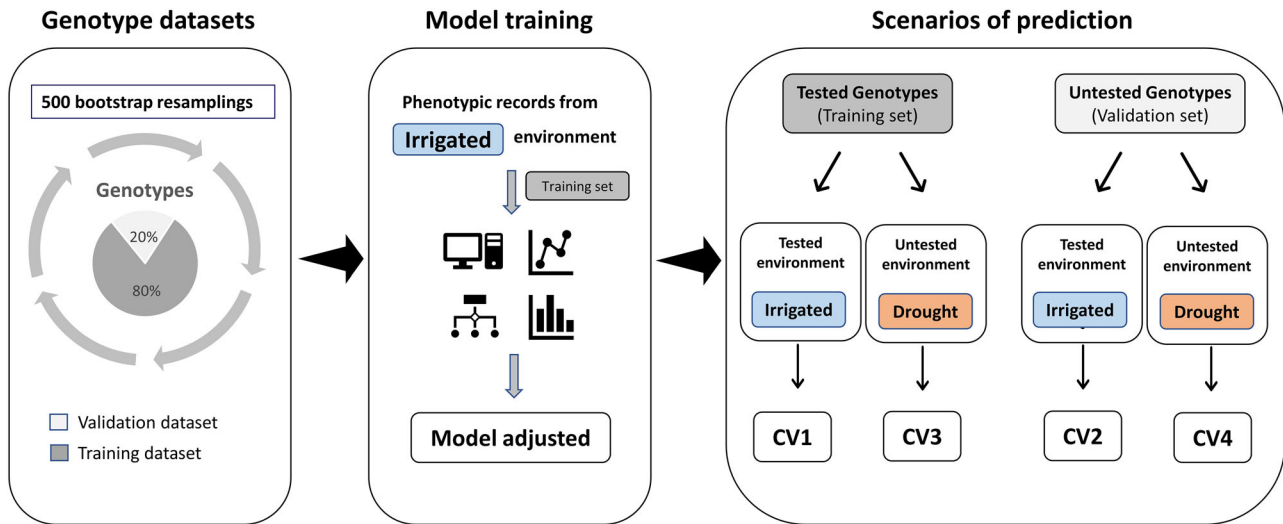
The variable importance scores for each regression model were obtained using the *caret::varImp* function. The acronym VI\_DAP represents a vegetation index (VI) calculated on specific flight dates (DAP). Once the model was trained, we used the *caret::predict* function with the corresponding dataset based on the prediction scenario. Each set of predictions was compared to the BLUPs of the ground phenotypic traits to calculate prediction accuracy, genotype rankings, and the root mean square error (RMSE) of prediction. We specifically chose the top 10 most important predictors from the model with the highest prediction accuracy scores in the CV2, CV3,



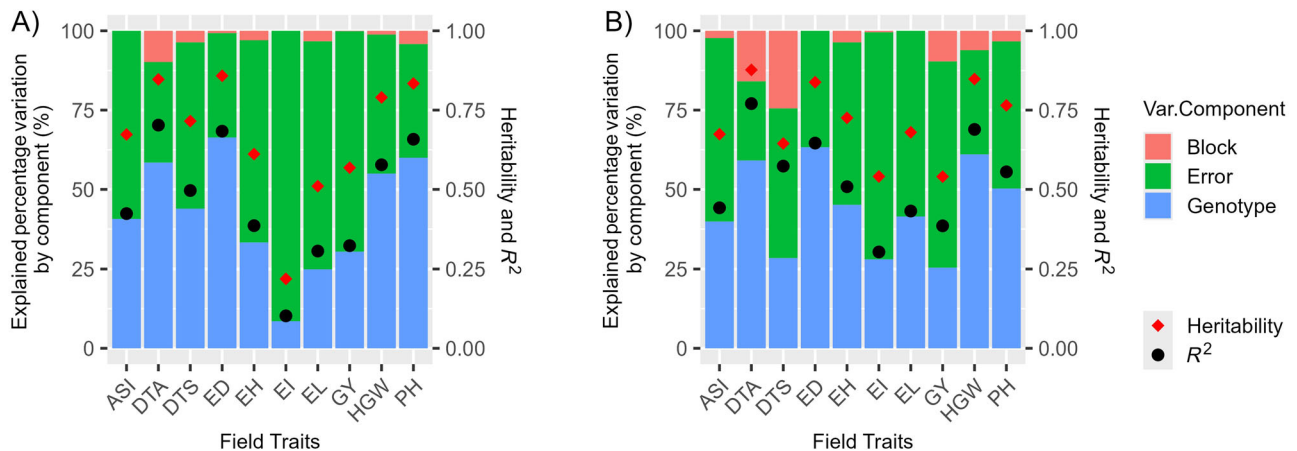
**FIGURE 2** Pipeline of imaging phenotyping process used in this work. BLUPs, best linear unbiased predictions; GCP, ground control points; RGB, red–green–blue.

and CV4 scenarios to identify the most influential vegetation indices and flight dates for the traits. The results of prediction efficiency, correlations involving the vegetation indices,

and key parameters from both phenotyping methods were presented in graphical visualizations using the *ggplot2* R package (Wickham, 2016).



**FIGURE 3** The main steps for prediction include how the genotypes are split in our training and validation datasets, how the model is trained with the available phenotypes from the environments under study, and further, the structure of the prediction scenarios, once we have a model already trained.



**FIGURE 4** Explained percentage variation by each component in Equation (3) for the ground phenotyping traits in the irrigated (A) and drought (B) trials. The left y axis represents the percentage of variation explained by the components, while the right y axis displays the heritability and  $R^2$  values. The traits included are anthesis silking interval (ASI), days to anthesis (DTA), days to silking (DTS), ear diameter (ED), ear height (EH), ear index (EI), ear length (EL), grain yield (GY), 100-grain weight (HGW), and plant height (PH).

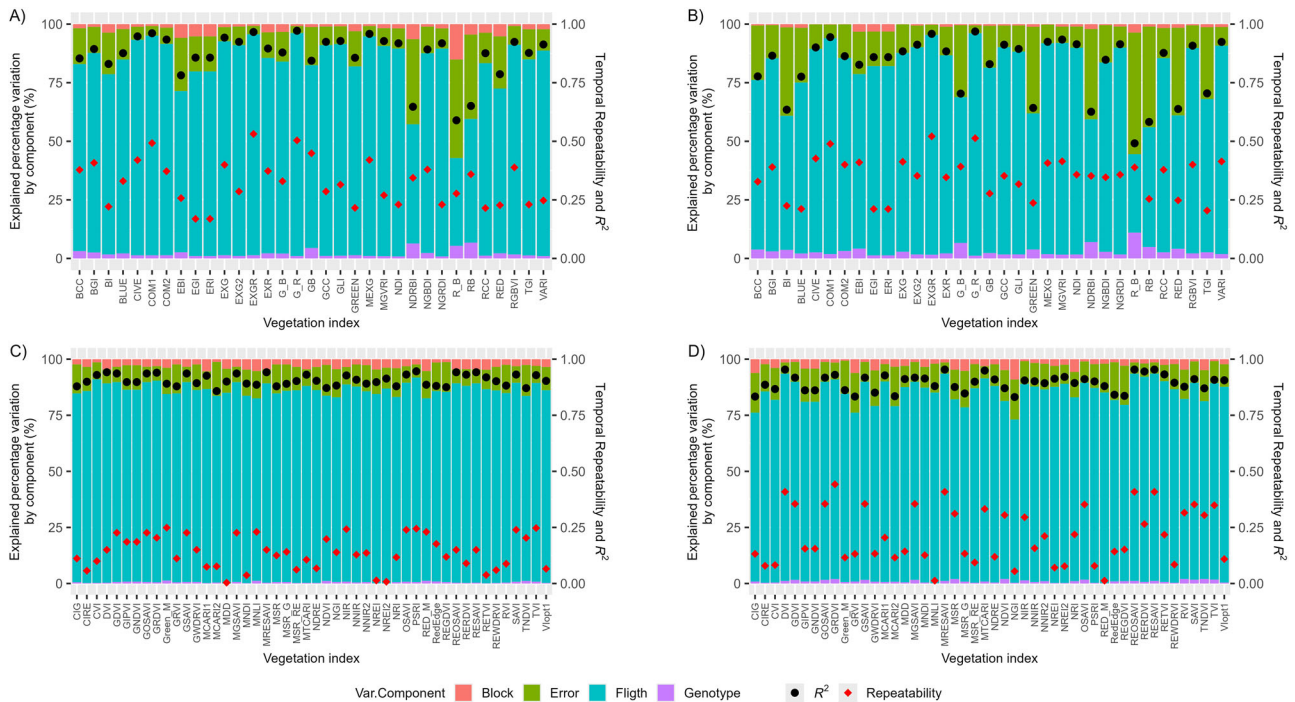
### 3 | RESULTS

#### 3.1 | Variances explained, heritability, temporal repeatability, and the sensitivity of the indices to drought

The two trials tested in this study showed similar results for  $R^2$  values (total variance explained by the model) and the heritability achieved for the traits (Figure 4). The only exception was observed for EI in the irrigated trial, where there was the least difference in  $R^2$  and heritability. For the other traits, the differences between both trials were small. In irrigated conditions, the heritability ranged from 0.22 (EI) to 0.86 (ED), and the  $R^2$  ranged from 0.10 (EI) to 0.70 (DTA), while in drought

conditions, these values were between 0.54 (EI) and 0.88 (DTA) for heritability, and 0.30 (EI) and 0.77 (DTA) for  $R^2$ . The genotype component, on average, explained almost the same variation in irrigated (42%) and drought (44%) conditions. Despite the similarities between both trials concerning the variances explained by each component and the calculated ratios ( $R^2$  and heritability), they yielded differently. On average, the irrigated trial yielded 9857 kg/ha, and the drought trial yielded 5624 kg/ha. The graph of grain yield density in each trial can be viewed in Figure S1.

The flight component in the nested design, for both sensors and trials, explained the most significant percentage of variation for all vegetation indices (Figure 5). Over the crop cycle and in both irrigated and drought conditions, there



**FIGURE 5** Explained percentage variation by each component in the nested design by Equation (1) for each vegetation index for the RGB sensor, irrigated (A) and drought (B) trial, and multispectral sensor, irrigated (C) and drought (D) trial. The left y axis corresponds to the explained percentage variation of the components, while the right y axis shows the temporal repeatability and  $R^2$  values.

was a significant variation in the temporal breeding values of the vegetation indices (Figure S2). Correlations within each phenomic dataset, for both irrigated and drought conditions, showed a wide range of variation, from  $-1$  to  $1$ , and larger blocks of strongly correlated vegetation indices were not observed (general mean value of  $0.08$ ), especially in the RGB dataset (Figure S3).

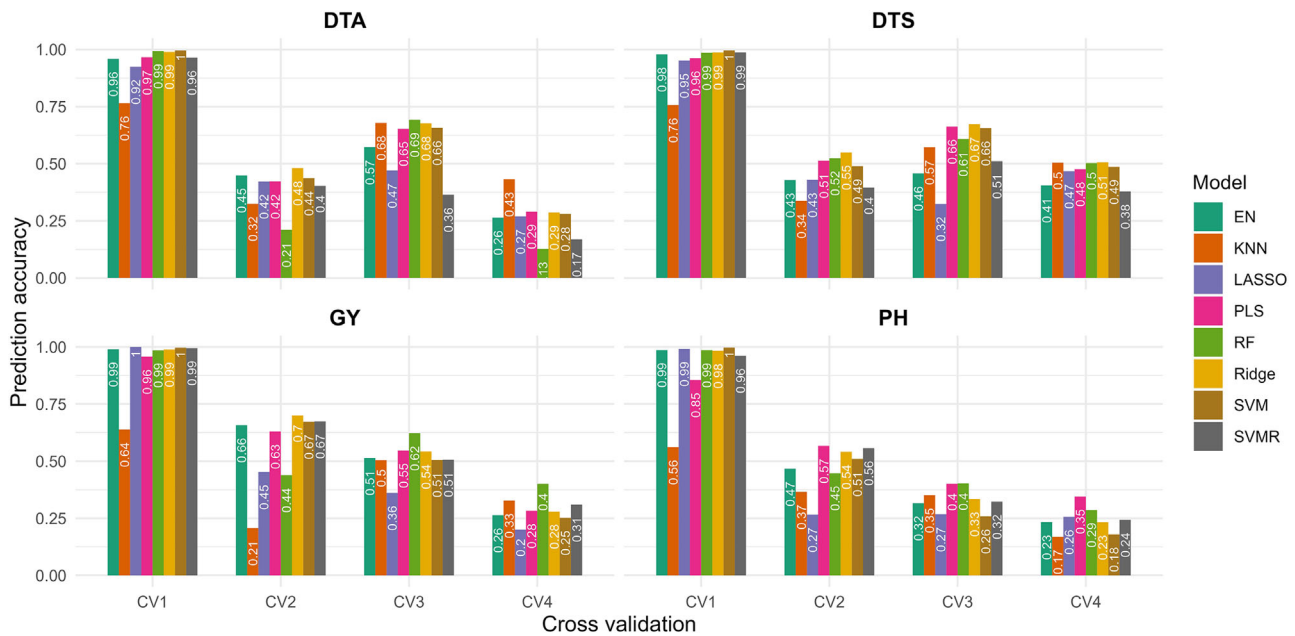
The flight component accounted for up to 95% of the variation in RGB phenomic data for the G\_R index while explaining the lowest percentage for R\_B (33% and 37% for drought and irrigated conditions, respectively). The multispectral phenomic data that were most influenced by the flight component were the PSRI index (91% for irrigated conditions) and the RESAVI index (92% for drought conditions), while the least affected were the single band red (81% for irrigated conditions) and the NGI index (73% for drought conditions).

The temporal repeatability of the RGB phenomic data was mostly concentrated in the range of  $0.25$ – $0.50$  (with an average of  $0.32$ ) and generally higher than the temporal repeatability of the multispectral data (with an average of  $0.16$ ), especially in the irrigated trial. The  $R^2$  value for the imaging phenotyping data was higher (with an average of  $0.86$ ) than the  $R^2$  observed for the ground phenotyping traits (with an average of  $0.50$ ). There was a slight difference in the average  $R^2$  values between multispectral and RGB phenomic data, with values of  $0.90$  and  $0.83$ , respectively.

Table S4 shows the changes in vegetation indices between the drought and irrigated trials during the vegetative stage. In the drought trial, the RGB vegetation indices R\_B, BI, RCC, and single band blue experienced significant reductions of over 100%. Conversely, VARI, normalized green–red difference index (NGRDI), NDI, COM2, and G\_B showed substantial increases in the drought trial. MDD and MNDI had the biggest decreases among the multispectral vegetation indices, while NREI and NREI2 saw the greatest increases. The multispectral indices showed more absolute value and range variation than the RGB indices.

### 3.2 | Prediction of phenotypic traits using high-resolution UAS imaging and temporal data

We focused our analysis on the most commonly used traits: DTA, DTS, GY, and PH. The predictions for the remaining six traits can be found in the Supporting Information. The phenomic predictions using RGB data (Figure 6) performed better than those using multispectral data (Figure S4). The predictions for DTS were very similar using both sensors ( $0.61$  for RGB and  $0.62$  for multispectral). However, the multispectral predictions reduced the accuracy by 12%, 47%, and 39% on average for DTA, GY, and PH, respectively. Therefore, all results henceforth will be based on RGB predictions.



**FIGURE 6** Red–green–blue (RGB) phenomic prediction accuracy (y axis) obtained by each model in four cross-validation scenarios of prediction (x axis) for the traits days to anthesis (DTA), days to silking (DTS), grain yield (GY), and plant height (PH). CV1 and CV2 indicate the prediction of tested and untested genotypes in the observed environment (irrigated), respectively. CV3 and CV4 indicate the prediction of tested and untested genotypes in the unobserved environment (drought), respectively. The models are elastic net (EN),  $k$ -nearest neighbor (KNN), least absolute shrinkage operator (LASSO), partial least squares (PLS), random forest (RF), ridge regression (Ridge), support vector machine (SVM), and support vector machine with radial kernel (SVMR).

The prediction accuracy for the other traits can be found in Table S5.

The KNN model consistently underperformed compared to other machine learning models when predicting the tested genotypes under irrigated conditions (CV1) for all traits. The other models performed similarly in this prediction scenario, except for PLS, which showed some deviation in predicting PH. However, the KNN model stood out when predicting untested genotypes for the trait DTA under drought conditions (CV4). For predicting tested genotypes under drought conditions (CV3) for the same trait, the KNN model performed similarly to Ridge, SVM, PLS, and RF and achieved the highest score. Ridge was the best predictor for untested genotypes for DTA in irrigated condition (CV2).

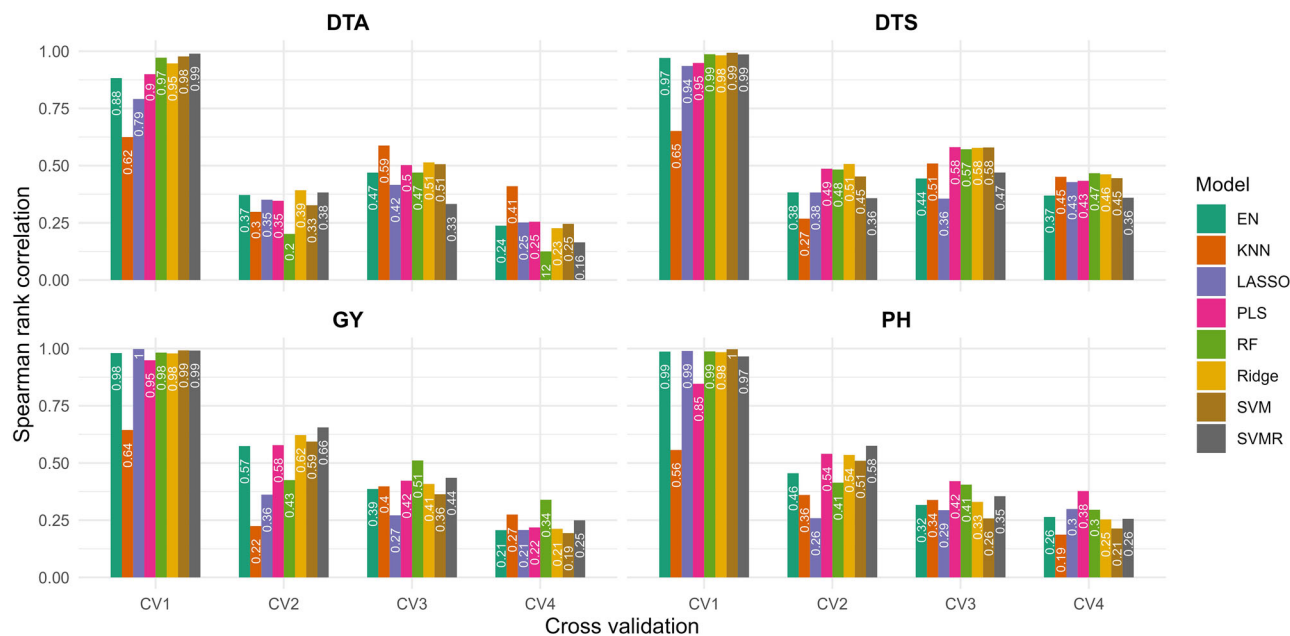
The best prediction performance for DTS was found using the Ridge model, both in irrigated and drought conditions. For GY, the RF model stood out for predicting drought conditions (CV3 and CV4), while it did not perform well in irrigated conditions (CV2). In this condition, the Ridge model was the best but very similar to the SVM, support vector machine with radial kernel (SVMR), and EN models. For the trait PH, the PLS model showed the highest predictions. In drought conditions, its performance was close to the RF model, and in irrigated conditions, for untested genotypes (CV2), it was close to the SVMR and Ridge models. As expected, there was a general trend of reduction in prediction accuracy from CV1 to CV4, mainly for GY and PH traits. Similarly, the RMSE

of prediction increased from CV1 to CV4 for all traits and models (Figures S5–S8 for the traits DTA, DTS, GY, and PH, respectively).

The genotypes were ranked based on a similar trend as the visualized prediction performance regarding machine learning models, prediction scenarios, and traits. Consequently, the best models for prediction accuracy were typically the best for ranking (Figure 7).

### 3.3 | Variable importance of temporal predictors

In the following results, the variables of importance included the vegetation index (VI) and the flight date (DAP) (the acronym VI\_DAP). PH had the weakest correlation with the vegetation indices, with no negative correlation (Figure 8A). The strongest correlations were found for the predictors EXGR\_95 (0.67), G\_R\_95 (0.60), BCC\_26 (0.50), and RCC\_119 (0.50). GY was mainly positively correlated with the predictors, as only three showed a negative correlation (CIVE\_99, BCC\_40, and BGI\_40). The strongest correlations were found for G\_B\_99 (0.72), R\_B\_99 (0.66), RCC\_95 (0.61), and CIVE\_99 (−0.73). These correlations cover the phenological stage of grain filling in our trials. It's important to note that among the common predictors between GY and flowering traits (DTA and DTS), the direction of correlation



**FIGURE 7** Red–green–blue (RGB) phenomic Spearman rank correlations (y axis) obtained by each model in four cross-validation scenarios of prediction (x axis) for the traits days to anthesis (DTA), days to silking (DTS), grain yield (GY), and plant height (PH). CV1 and CV2 indicate the prediction of tested and untested genotypes in the observed environment (irrigated), respectively. CV3 and CV4 indicate the prediction of tested and untested genotypes in the unobserved environments (drought), respectively. The models are elastic net (EN), *k*-nearest neighbor (KNN), least absolute shrinkage operator (LASSO), partial least squares (PLS), random forest (RF), ridge regression (Ridge), support vector machine (SVM), and support vector machine with radial kernel (SVMR).

(positive or negative) was always inverted. DTA and DTS had more negative correlations with the predictors. The strongest correlations for DTS were found with BCC\_40 (0.60), BGI\_40 (0.58), RCC\_95 (−0.71), and RCC\_99 (−0.65). For DTA, the strongest correlations were found with VARI\_91 (0.59), G\_R\_95 (0.52), RCC\_99 (−0.77), and RCC\_109 (−0.71). Among the common predictors between DTS and DTA, there was no discrepancy in the direction of correlation.

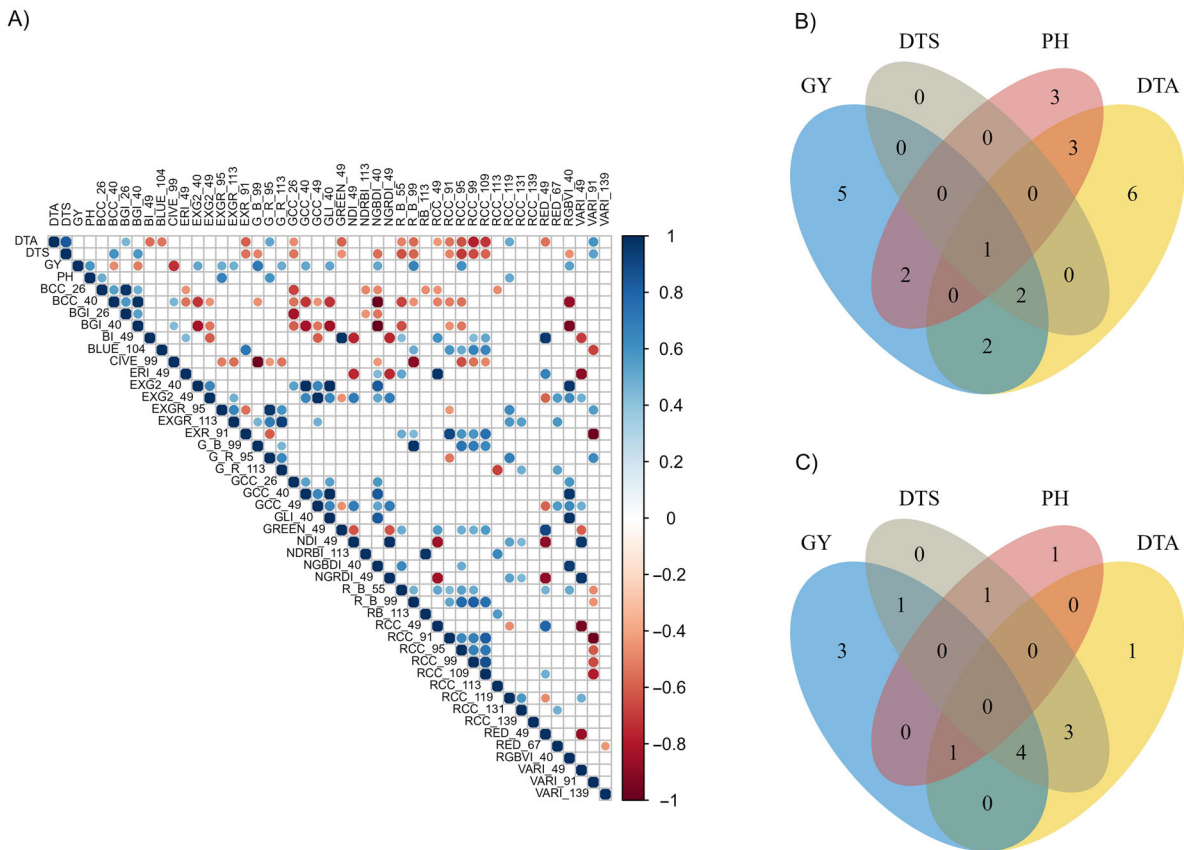
The investigation of the most important vegetation indices for prediction (excluding CV1 in all CV scenarios) revealed that the index GCC was common among all the traits (Figure 8B). Among GY, DTS, and DTA, we found two important indexes in common, RCC and NGRDI. No other index was commonly important in predicting any combination of three traits. The indices commonly important for predicting GY and PH were EXGR and G\_R, and for GY and DTA, these indices were the red band and VARI. The indices BGI, BCC, and EXG2 were common between PH and DTA. We found five, zero, three, and six indices as uniquely important when predicting GY, DTS, PH, and DTA, respectively. The most important indices for prediction, as represented by the Venn diagram, can be seen in Table S6.

The analysis of the most critical flight dates for prediction (in all CV scenarios, except CV1) revealed that there was no common date for all four traits (Figure 8C). However, 113 DAP was identified as an important flight date for predicting GY, DTA, and PH. Flight dates 91, 119, 131, and 139 were

equally important in predicting GY, DTA, and DTS. 95 DAP stood out for the prediction of DTS and PH, while 99 DAP was highlighted for the prediction of GY and DTS. Among the flowering traits, DTA and DTS, the dates 26, 49, and 109 DAP were commonly important. We also identified three, zero, one, and one flight dates that were important only to predict GY, DTS, PH, and DTA separately. The most important flight dates for prediction, as shown in the Venn diagram, can be found in Table S6.

## 4 | DISCUSSION

Using UASs for field-based phenotyping marks a significant advancement in agricultural research. By harnessing the speed and precision of UAS technology, researchers can efficiently and accurately conduct trial evaluations. This approach allows for rapidly assessing various crop traits, revolutionizing how agricultural data are measured and analyzed. As plant response to water stress involves physiological and morphological changes over time, a robust evaluation method for assessing drought tolerance should incorporate temporal data responses. However, it's important to note that many studies exploring the application of UAS phenotyping in plant breeding have typically conducted evaluations at only limited and sporadic time points throughout the growth cycle (Aguate et al., 2017; Jackson et al., 2023; Winn et al., 2023). Our



**FIGURE 8** The most important predictors, vegetation indices and flight dates, and their correlations. (A) Significant correlations among the most important temporal predictors and the traits for the prediction in drought condition (only CV3 and CV4 scenarios). The size and color of the circles indicate the strength of the correlation, where blue and red mean positive and negative significant correlations, respectively. (B) Venn diagram for the most important vegetation indices for prediction (all scenarios, except CV1). (C) Venn diagram for the most important flight dates (days after planting [DAP]) for prediction (all scenarios, except CV1). The variables importance (top 10 most important) came from the models with the highest scores for prediction in each scenario, that is, ridge (CV2), random forest (CV3), and KNN (CV4) for DTA; ridge (CV2, CV3, and CV4) for DTS; ridge (CV2) and random forest (CV3 and CV4) for GY; and partial least squares (CV2, CV3, and CV4) for PH. DTA, DTS, GY, and PH mean days to anthesis, days to silking, grain yield, and plant height, respectively.

findings reveal that the variation of the vegetation indices captured by the flight time in the nested design had been typically ignored (Figure 5).

The optimal frequency of flights for agricultural monitoring remains uncertain and should be determined based on the specific crop and trait being studied. For example, Volpato et al. (2021) observed that conducting flights every 2 weeks was less effective in predicting soybean (*Glycine max* L.) maturity compared to flights conducted weekly. Similarly, Wu et al. (2019) emphasized the importance of conducting flights during the senescence period in maize to accurately characterize senescence progression, especially for predicting yield. Supporting the need for multiple data points, particularly in stress response studies, Figure S2 demonstrates the variation in vegetation indexes throughout the crop cycle.

The reliability of a vegetation index in capturing genotypic differences and environmental influences determines its practical usefulness. While some studies have reported high

repeatability values for vegetation indices under both optimal and stressed conditions, our research found somewhat lower repeatability values in comparison (Adak et al., 2021; Adak, Kang, et al., 2023; Adak, Murray, et al., 2023). Nonetheless, certain vegetation indices showed repeatability levels that closely matched the heritability of the trait measured using traditional ground evaluation methods. This is particularly promising, especially for assessing end-of-life traits like grain yield. The ability of these vegetation indices to provide consistent and reliable measurements, capture the variation along the flights, change according to the environment, and provide a faster way to get phenotypes holds potential for early and repeated assessments throughout the crop cycle. Thus enriching the available dataset for decision-making and deepening our understanding of the biological basis of crop stress responses.

It is crucial to consider different vegetation indices, as the most appropriate one will depend on the specific trait,

prediction model, and environmental factors involved (Adak, Murray, et al., 2023; Aguete et al., 2017; DeSalvio et al., 2022; Dodig et al., 2019). Additionally, the data collection process is the same, and the vegetation indices contain distinct information, as depicted in Figure S3. This figure illustrates that the vegetation indices exhibit a low to moderate correlation between them and within the same index over time. The issue of multicollinearity does not seem to be a concern in phenomic surveys (Adak, Kang, et al., 2023).

In-field assessment of water stress can be pretty challenging. Having a tool that can quickly and accurately identify stress early would be advantageous. A vegetation index that can reliably indicate the water status of plants has the potential to be directly used in regular trial evaluations, as it allows fast data collection and enables near-real-time intervention. Our study has shown that RGB and multispectral vegetation indices hold promise for easily monitoring plant water status. This is because noticeable differences in drought conditions compared to irrigated conditions were observed for both indices, indicating their dynamic and responsive nature to stress (Dodig et al., 2019, 2021; Winterhalter et al., 2011). Such valuable information can help determine the right management practices for effective genotype screening.

#### 4.1 | Predicting phenotypes with machine learning at high spatial and temporal resolution

The RGB sensor performed better than the multispectral sensor in prediction accuracies. The results of Adak, Murray et al. (2023) corroborate this finding for the same traits and CV scenarios studied here. The multispectral predictions reported by these authors were 32%, 16%, 1%, and 26% lower than the RGB predictions for DTA, DTS, GY, and PH, respectively. The lower resolution and fewer flights for the multispectral predictions could have impaired their performance since all the remaining factors were the same for both sensors. As a result, a cost-effective and straightforward sensor is sufficient for predicting essential traits in maize. This discovery shows promise for academic institutions, public initiatives, and low-resource programs. Therefore, it may not be advisable at this stage to invest in additional spectral bands at a higher sensor cost for prediction purposes.

As in genomic selection, phenomic prediction is influenced by both the trait being studied and the model used (Adak, Murray, et al., 2023; DeSalvio et al., 2022; Zhu et al., 2021, 2022). While no single model consistently stood out for general use, certain models did show exceptional prediction accuracy for specific scenarios and traits. We can highlight the models ridge, RF, and KNN for DTA; ridge for DTS; ridge and RF for GY; and PLS for PH.

In this work, we found that the Ridge and RF models consistently showed the highest stability for prediction across

all traits in most scenarios. In particular, Ridge proved to be the best model for predicting the grain yield (GY) of maize hybrids in the CV2 scenario, with our study achieving a higher prediction accuracy (0.70) compared to Adak et al. (2021) (0.60). However, our study's best models for flowering traits (ridge and KNN) differed from those they identified (lasso and elastic net). Additionally, our prediction accuracy for grain yield in the CV3 scenario using the RF model (0.62) was similar to the results reported by Adak, Murray et al. (2023) with the ridge model (0.59). However, in the CV4 scenario, our results (0.40) were lower than those using the same models (0.57). For the trait PH, our prediction accuracies were generally lower than those reported by these authors (0.57, 0.40, and 0.35 for CV2, CV3, and CV4, respectively). This trend was also observed for flowering traits.

The accuracy of our predictions for DTA, DTS, and PH in the CV3 scenario was higher than that reported by Adak, Kang et al. (2023). Their study, which utilized a population of maize recombinant inbred lines (RILs) and phenomic data from RGB or multispectral sensors, did not use a nested model as we did. However, in the CV4 scenario, their predictions were similar to ours. Chatterjee et al. (2023) obtained prediction accuracy for the GY of maize hybrids comparable to ours. They also noted that the temporal vegetation indices were better predictors than temporal canopy height measurements or cumulative vegetation indices over the growing cycle. Our prediction accuracy for GY (ranging from 0.40 to 0.70 across CV scenarios) matched the results reported by Wu et al. (2019) (0.40–0.69). That study also considered several flights throughout the growth cycle, with a similar number of genotypes under evaluation (25). Still, it included more years (3) and employed PLS regression on principal components.

Our findings showed that temporal phenomic prediction can be a valuable tool for plant breeding, even in small-scale programs, with high spatial and temporal resolution. Our achieved predictions are comparable to those reported in the literature despite our smaller experimental scale. It is expected that better results can be achieved in larger populations through more robust model training. Therefore, temporal phenomic prediction can be widely applied in commercial breeding programs, enhancing the efficiency and accuracy of plant breeding efforts.

#### 4.2 | Variable importance for prediction and correlations

In drought conditions, certain combinations of vegetation indices and flight dates showed strong correlations with the traits in spot assessments (Figure 8A). These indices can guide further studies and deserve attention. Previous works indicate that they can be used independently to infer the respective trait (Adak et al., 2021; Adak, Murray, et al., 2023; Aguete

et al., 2017). Furthermore, they can help identify the optimal time to assess specific plant health conditions, as each vegetation index dynamically reflects a particular physiological status (Dodig et al., 2019, 2021; Winterhalter et al., 2011).

Another important aspect to consider is the consistency in the correlation between vegetation indices and traits and the correlations among the traits themselves. In our investigation, grain yield (GY) was negatively correlated with flowering traits (DTA and DTS). The exact negative correlation was observed for the most significant vegetation indices associated with these traits. Therefore, besides providing valuable information about environmental conditions and plant response, vegetation indices also elucidate the relationships among the studied traits.

Several indices, including NGRDI, RCC, and VARI, stood out in our study and have also been highlighted in other works on predicting phenotypes in maize. We recommend that these indices be incorporated into routine evaluations in maize trials. Adak et al. (2021) also identified NGRDI and VARI as among the most important indices for predicting GY in maize. Similarly, Adak, Murray et al. (2023) found the RCC index to be a key predictor for the traits evaluated here. Chatterjee et al. (2023), in their investigation of prediction models in maize and exploration of different ways to incorporate vegetation indices, consistently found NGRDI to be the best predictor of GY in models considering vegetation indices at the temporal level, similar to our approach. Furthermore, RCC was identified as the most important index for predicting rust and senescence in maize hybrids, and it strongly correlated with GY, as shown in the work by DeSalvio et al. (2022). Apart from their predictive importance, the RGB indices NGRDI, RCC, and VARI were particularly suited for indicating drought response, as they varied markedly between the trials.

The vegetative stage encompassed several significant flight dates for predicting key traits, which is consistent with findings in other studies on maize grown in stressful conditions (Adak et al., 2021; Aguete et al., 2017; Chatterjee et al., 2023). This suggests that end-of-life traits, typically assessed only once using traditional methods, can benefit from UAS phenotyping (Adak et al., 2021; Adak, Murray, et al., 2023). UAS phenotyping allows for more efficient monitoring and study of these traits during selection, providing a cost-effective and quicker alternative to manual measurements.

The results presented in this work showed that phenomic prediction has proven to be a promising approach. We were able to predict important traits satisfactorily, indicating the potential for routine application of UAS phenotyping in field evaluations, which could save both time and labor. It is crucial that a phenotyping tool intended for use in transgenic trials for repeated evaluations, such as UAS phenotyping, remain cost effective, as regulatory requirements already

significantly increase the burden on this type of research (Gómez-Galera et al., 2012; Yassitepe et al., 2021). Another advantage is the substantial amount of information gathered through repeated evaluations over the crop cycle, leading to more reliable results, especially under abiotic stress conditions. Among the range of omics tools available for prediction, phenomics stands out as the simplest, most flexible, and most accessible method for measuring secondary-related traits. We have demonstrated its utility using inexpensive equipment and simple and freely available tools shared with the community, which can help promote phenomic predictions among low-resource programs and institutions. This, in turn, can enhance the design and allocation of genotypes within trial networks.

## AUTHOR CONTRIBUTIONS

**Helcio Duarte Pereira:** Conceptualization; data curation; formal analysis; methodology; writing—original draft. **Juliana Vieira Almeida Nonato:** Data curation; investigation; methodology; validation. **Rafaela Caroline Rangni Molto-caro Duarte:** Data curation; investigation; methodology; validation. **Isabel Rodrigues Gerhardt:** Funding acquisition; project administration; resources; writing—review and editing. **Ricardo Augusto Dante:** Funding acquisition; project administration; resources; writing—review and editing. **Paulo Arruda:** Conceptualization; funding acquisition; project administration; supervision; writing—review and editing. **Juliana Erika de Carvalho Teixeira Yassitepe:** Conceptualization; data curation; funding acquisition; methodology; project administration; resources; supervision; validation; visualization; writing—review and editing.

## ACKNOWLEDGMENTS

This work was funded by grant 2016/23218-0 “Genomics for Climate Change Research Center (GCCRC)” from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP). Helcio Duarte Pereira and Juliana Vieira Almeida Nonato received a technical training fellowship from FAPESP (2023/11640-2 and 2022/04930-1, respectively). This study was partly funded by Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA) and Universidade Estadual de Campinas (UNICAMP).

## CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.


## DATA AVAILABILITY STATEMENT

A dataset containing an example of an orthomosaic (TIF format) and its corresponding polygons of shape file (from QGIS software) used to clip the plots and calculate the RGB vegetation indices was given in the folder “Extracting vegetation indices.” Phenotypic records for two traits (plant height and grain yield) and for all the 35 RGB vegetation indices used

in this work to perform phenomic prediction for all the eight models and the four cross-validation scenarios were given as examples in the folder “Phenomic prediction.” All the codes and the datasets described are available at <https://www.gccrc.unicamp.br/publications/> and <https://doi.org/10.5061/dryad.0zpc8677b>.

## ORCID

Helcio Duarte Pereira  <https://orcid.org/0000-0002-2837-9396>

Juliana Vieira Almeida Nonato  <https://orcid.org/0000-0003-4448-4652>

Rafaela Caroline Rangni Moltocaro Duarte  <https://orcid.org/0000-0003-2622-3758>

Isabel Rodrigues Gerhardt  <https://orcid.org/0000-0003-1397-0199>

Ricardo Augusto Dante  <https://orcid.org/0000-0002-2738-4433>

Paulo Arruda  <https://orcid.org/0000-0002-8365-731X>

Juliana Erika de Carvalho Teixeira Yassitepe  <https://orcid.org/0000-0002-6862-3031>

## REFERENCES

- Adak, A., Kang, M., Anderson, S. L., Murray, S. C., Jarquin, D., Wong, R. K., & Katzfuß, M. (2023). Phenomic data-driven biological prediction of maize through field-based high-throughput phenotyping integration with genomic data. *Journal of Experimental Botany*, *74*(17), 5307–5326. <https://doi.org/10.1093/jxb/erad216>
- Adak, A., Murray, S. C., & Anderson, S. L. (2023). Temporal phenomic predictions from unoccupied aerial systems can outperform genomic predictions. *G3 Genes|Genomes|Genetics*, *13*(1), jkac294. <https://doi.org/10.1093/g3journal/jkac294>
- Adak, A., Murray, S. C., Božinović, S., Lindsey, R., Nakasagga, S., Chatterjee, S., Anderson, S. L., & Wilde, S. (2021). Temporal vegetation indices and plant height from remotely sensed imagery can predict grain yield and flowering time breeding value in maize via machine learning regression. *Remote Sensing*, *13*(11), 2141. <https://doi.org/10.3390/rs13112141>
- Aguate, F. M., Trachsel, S., Pérez, L. G., Burgueño, J., Crossa, J., Balzarini, M., Gouache, D., Bogard, M., & de los Campos, G. (2017). Use of hyperspectral image data outperforms vegetation indices in prediction of maize yield. *Crop Science*, *57*(5), 2517–2524. <https://doi.org/10.2135/cropsci2017.01.0007>
- Anderson, S. L., Murray, S. C., Malambo, L., Ratcliff, C., Popescu, S., Cope, D., Chang, A., Jung, J., & Thomasson, J. A. (2019). Prediction of maize grain yield before maturity using improved temporal height estimates of unmanned aerial systems. *The Plant Phenome Journal*, *2*(1), 1–15. <https://doi.org/10.2135/tppj2019.02.0004>
- Araus, J. L., & Cairns, J. E. (2014). Field high-throughput phenotyping: The new crop breeding frontier. *Trends in plant science*, *19*(1), 52–61. <https://doi.org/10.1016/j.tplants.2013.09.008>
- Chafai, N., Hayah, I., Houaga, I., & Badaoui, B. (2023). A review of machine learning models applied to genomic prediction in animal breeding. *Frontiers in genetics*, *14*, 1150596. <https://doi.org/10.3389/fgene.2023.1150596>
- Chatterjee, S., Adak, A., Wilde, S., Nakasagga, S., & Murray, S. C. (2023). Cumulative temporal vegetation indices from unoccupied aerial systems allow maize (*Zea mays* L.) hybrid yield to be estimated across environments with fewer flights. *PLoS One*, *18*(1), e0277804. <https://doi.org/10.1371/journal.pone.0277804>
- Covarrubias-Pazarán, G. (2016). Genome-assisted prediction of quantitative traits using the R package sommer. *PLoS One*, *11*(6), e0156744. <https://doi.org/10.1371/journal.pone.0156744>
- DeSalvio, A. J., Adak, A., Murray, S. C., Wilde, S. C., & Isakeit, T. (2022). Phenomic data-facilitated rust and senescence prediction in maize using machine learning algorithms. *Scientific Reports*, *12*(1), Article 7571. <https://doi.org/10.1038/s41598-022-11591-0>
- Dodig, D., Božinović, S., Nikolić, A., Zorić, M., Vančetović, J., Ignjatović-Mičić, D., Delić, N., Weigelt-Fischer, K., Altmann, T., & Junker, A. (2021). Dynamics of maize vegetative growth and drought adaptability using image-based phenotyping under controlled conditions. *Frontiers in Plant Science*, *12*, 652116. <https://doi.org/10.3389/fpls.2021.652116>
- Dodig, D., Božinović, S., Nikolić, A., Zorić, M., Vančetović, J., Ignjatović-Mičić, D., Delić, N., Weigelt-Fischer, K., Junker, A., & Altmann, T. (2019). Image-derived traits related to mid-season growth performance of maize under nitrogen and water stress. *Frontiers in Plant Science*, *10*, 434112. <https://doi.org/10.3389/fpls.2019.00814>
- Erenstein, O., Jaleta, M., Sonder, K., Mottaleb, K., & Prasanna, B. M. (2022). Global maize production, consumption and trade: Trends and R&D implications. *Food Security*, *14*(5), 1295–1319. <https://doi.org/10.1007/s12571-022-01288-7>
- FAOSTAT. (2022). *Crops and livestock products*. <https://www.fao.org/faostat/en/#data/QCL>
- Gómez-Galera, S., Twyman, R. M., Sparrow, P. A., Van Droogenbroeck, B., Custers, R., Capell, T., & Christou, P. (2012). Field trials and tribulations—Making sense of the regulations for experimental field trials of transgenic crops in Europe. *Plant Biotechnology Journal*, *10*(5), 511–523. <https://doi.org/10.1111/j.1467-7652.2012.00681.x>
- Hijmans, R. (2024). *terra: Spatial data analysis* (R package version 1.7-77) [Computer software]. <https://CRAN.R-project.org/package=terra>, <https://rspatial.org/>
- Jackson, R., Buntjer, J. B., Bentley, A. R., Lage, J., Byrne, E., Burt, C., Jack, P., Berry, S., Flatman, E., Poupard, B., Smith, S., Hayes, C., Barber, T., Love, B., Gaynor, R. C., Gorjanc, G., Howell, P., Mackay, I. J., Hickey, J. M., & Ober, E. S. (2023). Phenomic and genomic prediction of yield on multiple locations in winter wheat. *Frontiers in Genetics*, *14*, 1164935. <https://doi.org/10.3389/fgene.2023.1164935>
- Jarquin, D., Howard, R., Liang, Z., Gupta, S. K., Schnable, J. C., & Crossa, J. (2020). Enhancing hybrid prediction in pearl millet using genomic and/or multi-environment phenotypic information of inbreds. *Frontiers in Genetics*, *10*, 1294. <https://doi.org/10.3389/fgene.2019.01294>
- Jarquin, D., Lemes da Silva, C., Gaynor, R. C., Poland, J., Fritz, A., Howard, R., Battenfield, S., & Crossa, J. (2017). Increasing genomic-enabled prediction accuracy by modeling genotype × environment interactions in Kansas wheat. *The Plant Genome*, *10*(2). <https://doi.org/10.3835/plantgenome2016.12.0130>
- Krause, M. R., Mondal, S., Crossa, J., Singh, R. P., Pinto, F., Haghhighattalab, A., Shrestha, S., Rutkoski, J., Gore, M. A., Sorrells, M. E., & Poland, J. (2020). Aerial high-throughput phenotyping enables indirect selection for grain yield at the early generation, seed-

- limited stages in breeding programs. *Crop Science*, 60(6), 3096–3114. <https://doi.org/10.1002/csc2.20259>
- Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, 28(5), 1–26. <https://doi.org/10.18637/jss.v028.i05>
- Lane, H. M., Murray, S. C., Montesinos-López, O. A., Montesinos-López, A., Crossa, J., Rooney, D. K., Barrero-Farfan, I. D., De La Fuente, G. N., & Morgan, C. L. (2020). Phenomic selection and prediction of maize grain yield from near-infrared reflectance spectroscopy of kernels. *The Plant Phenome Journal*, 3(1), e20002. <https://doi.org/10.1002/ppj2.20002>
- Li, Y., Zhang, P., Sheng, W., Zhang, Z., Rose, R. J., & Song, Y. (2023). Securing maize reproductive success under drought stress by harnessing CO<sub>2</sub> fertilization for greater productivity. *Frontiers in Plant Science*, 14, 1221095. <https://doi.org/10.3389/fpls.2023.1221095>
- Lobell, D. B., Hammer, G. L., McLean, G., Messina, C., Roberts, M. J., & Schlenker, W. (2013). The critical role of extreme heat for maize production in the United States. *Nature Climate Change*, 3(5), 497–501. <https://doi.org/10.1038/nclimate1832>
- Lobell, D. B., Schlenker, W., & Costa-Roberts, J. (2011). Climate trends and global crop production since 1980. *Science*, 333(6042), 616–620. <https://doi.org/10.1126/science.1204531>
- Matias, F. I., Caraza-Harter, M. V., & Endelman, J. B. (2020). FIELDImageR: An R package to analyze orthomosaic images from agricultural field trials. *The Plant Phenome Journal*, 3(1), e20005. <https://doi.org/10.1002/ppj2.20005>
- McMillen, M. S., Mahama, A. A., Sibiya, J., Lübberstedt, T., & Suza, W. P. (2022). Improving drought tolerance in maize: Tools and techniques. *Frontiers in Genetics*, 13, 1001001. <https://doi.org/10.3389/fgene.2022.1001001>
- Messina, C., McDonald, D., Poffenbarger, H., Clark, R., Salinas, A., Fang, Y., Gho, C., Tang, T., Graham, G., Hammer, G. L., & Cooper, M. (2021). Reproductive resilience but not root architecture underpins yield improvement under drought in maize. *Journal of Experimental Botany*, 72(14), 5235–5245. <https://doi.org/10.1093/jxb/erab231>
- Montesinos-López, O. A., Montesinos-López, A., Crossa, J., de Los Campos, G., Alvarado, G., Suchismita, M., Rutkoski, J., González-Pérez, L., & Burgueño, J. (2017). Predicting grain yield using canopy hyperspectral reflectance in wheat breeding data. *Plant Methods*, 13, Article 4. <https://doi.org/10.1186/s13007-016-0154-2>
- Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R<sup>2</sup> from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4(2), 133–142. <https://doi.org/10.1111/j.2041-210x.2012.00261.x>
- Pandey, M. K., Chaudhari, S., Jarquin, D., Janila, P., Crossa, J., Patil, S. C., Sundravandana, S., Khare, D., Bhat, R. S., Radhakrishnan, T., Hickey, J. M., & Varshney, R. K. (2020). Genome-based trait prediction in multi-environment breeding trials in groundnut. *Theoretical and Applied Genetics*, 133, 3101–3117. <https://doi.org/10.1007/s00122-020-03658-1>
- QGIS Development Team. (2024). *QGIS geographic information system* (Open Source Geospatial Foundation Project). <http://qgis.osgeo.org>
- Rincent, R., Charpentier, J. P., Faivre-Rampant, P., Paux, E., Le Gouis, J., Bastien, C., & Segura, V. (2018). Phenomic selection is a low-cost and high-throughput method based on indirect predictions: Proof of concept on wheat and poplar. *G3 Genes|Genomes|Genetics*, 8(12), 3961–3972. <https://doi.org/10.1534/g3.118.200760>
- Robert, P., Auzanneau, J., Goudemand, E., Oury, F. X., Rolland, B., Heumez, E., Bouchet, S., Le Gouis, J., & Rincent, R. (2022). Phenomic selection in wheat breeding: Identification and optimisation of factors influencing prediction accuracy and comparison to genomic selection. *Theoretical and Applied Genetics*, 135, 895–914. <https://doi.org/10.1007/s00122-021-04005-8>
- Sandhu, K., Patil, S. S., Pumphrey, M., & Carter, A. (2021). Multitrait machine- and deep-learning models for genomic selection using spectral information in a wheat breeding program. *The Plant Genome*, 14(3), e20119. <https://doi.org/10.1002/tpg2.20119>
- Volpato, L., Dobbels, A., Borem, A., & Lorenz, A. J. (2021). Optimization of temporal UAS-based imagery analysis to estimate plant maturity date for soybean breeding. *The Plant Phenome Journal*, 4(1), e20018. <https://doi.org/10.1002/ppj2.20018>
- Wang, X., Müller, C., Elliot, J., Mueller, N. D., Ciaias, P., Jägermeyr, J., Gerber, J., Dumas, P., Wang, C., Yang, H., Li, L., Deryng, D., Folberth, C., Liu, W., Makowski, D., Olin, S., Pugh, T. A. M., Reddy, A., Schmid, E., ... Piao, S. (2021). Global irrigation contribution to wheat and maize yield. *Nature Communications*, 12, Article 1235. <https://doi.org/10.1038/s41467-021-21498-5>
- Weiß, T. M., Zhu, X., Leiser, W. L., Li, D., Liu, W., Schipprack, W., Melchinger, A. E., Hahn, V., & Würschum, T. (2022). Unraveling the potential of phenomic selection within and among diverse breeding material of maize (*Zea mays* L.). *G3 Genes|Genomes|Genetics*, 12(3), jkab445. <https://doi.org/10.1093/g3journal/jkab445>
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag. <https://ggplot2.tidyverse.org>
- Wijewardane, N. K., Zhang, H., Yang, J., Schnable, J. C., Schachtman, D. P., & Ge, Y. (2023). A leaf-level spectral library to support high-throughput plant phenotyping: Predictive accuracy and model transfer. *Journal of Experimental Botany*, 74(14), 4050–4062. <https://doi.org/10.1093/jxb/erad129>
- Winn, Z. J., Amsberry, A. L., Haley, S. D., DeWitt, N. D., & Mason, R. E. (2023). Phenomic versus genomic prediction—A comparison of prediction accuracies for grain yield in hard winter wheat lines. *The Plant Phenome Journal*, 6(1), e20084. <https://doi.org/10.1002/ppj2.20084>
- Winterhalter, L., Mistele, B., Jampatong, S., & Schmidhalter, U. (2011). High throughput phenotyping of canopy water mass and canopy temperature in well-watered and drought stressed tropical maize hybrids in the vegetative stage. *European Journal of Agronomy*, 35(1), 22–32. <https://doi.org/10.1016/j.eja.2011.03.004>
- Wu, G., Miller, N. D., De Leon, N., Kaeppler, S. M., & Spalding, E. P. (2019). Predicting *Zea mays* flowering time, yield, and kernel dimensions by analyzing aerial images. *Frontiers in Plant Science*, 10, 467351. <https://doi.org/10.3389/fpls.2019.01251>
- Yan, S., Weng, B., Jing, L., & Bi, W. (2023). Effects of drought stress on water content and biomass distribution in summer maize (*Zea mays* L.). *Frontiers in Plant Science*, 14, 1118131. <https://doi.org/10.3389/fpls.2023.1118131>
- Yassitepe, J. E. D. C. T., da Silva, V. C. H., Hernandez-Lopes, J., Dante, R. A., Gerhardt, I. R., Fernandes, F. R., da Silva, P. A., Vieira, L. R., Bonatti, V., & Arruda, P. (2021). Maize transformation: From plant material to the release of genetically modified and edited varieties. *Frontiers in Plant Science*, 12, 766702. <https://doi.org/10.3389/fpls.2021.766702>
- Yuan, X., Wang, Y., Ji, P., Wu, P., Sheffield, J., & Otkin, J. A. (2023). A global transition to flash droughts under climate change. *Science*, 380(6641), 187–191. <https://doi.org/10.1126/science.abn6301>
- Zheng, H., Bian, Q., Yin, Y., Ying, H., Yang, Q., & Cui, Z. (2018). Closing water productivity gaps to achieve food and water security

- for a global maize supply. *Scientific Reports*, 8(1), Article 14762. <https://doi.org/10.1038/s41598-018-32964-4>
- Zhu, X., Leiser, W. L., Hahn, V., & Würschum, T. (2021). Phenomic selection is competitive with genomic selection for breeding of complex traits. *The Plant Phenome Journal*, 4(1), e20027. <https://doi.org/10.1002/ppj2.20027>
- Zhu, X., Maurer, H. P., Jenz, M., Hahn, V., Ruckelshausen, A., Leiser, W. L., & Würschum, T. (2022). The performance of phenomic selection depends on the genetic architecture of the target trait. *Theoretical and Applied Genetics*, 135, 653–665. <https://doi.org/10.1007/s00122-021-03997-7>

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Pereira, H. D., Nonato, J. V. A., Duarte, R. C. R. M., Gerhardt, I. R., Dante, R. A., Arruda, P., & Yassitepe, J. E. d. C. T. (2025). Temporal field phenomics of transgenic maize events subjected to drought stress: Cross-validation scenarios and machine learning models. *The Plant Phenome Journal*, 8, e70015. <https://doi.org/10.1002/ppj2.70015>