Using UAV-based hyperspectral imaging and functional regression to assist in predicting grain yield and related traits in wheat under heat-related stress environments for the purpose of stable yielding genotypes

Lucas Costa¹ · Jordan McBreen² · Yiannis Ampatzidis¹ · Jia Guo² · Mostafa Reisi Gahrooei³ · Md Ali Babar²

Accepted: 4 September 2021 / Published online: 18 September 2021
© The Author(s) 2021

Abstract

Quantifying certain physiological traits under heat-stress is crucial for maximizing genetic gain for wheat yield and yield-related components. In-season estimation of different physiological traits related to heat stress tolerance can ensure the finding of germplasm, which could help in making effective genetic gains in yield. However, estimation of those complex traits is time- and labor-intensive. Unmanned aerial vehicle (UAV) based hyperspectral imaging could be a powerful tool to estimate indirectly in-season genetic variation for different complex physiological traits in plant breeding that could improve genetic gains for different important economic traits, like grain yield. This study aims to predict in-season genetic variations for cellular membrane thermostability (CMT), yield and yield related traits based on spectral data collected from UAVs; particularly, in cases where there is a small sample size to collect data from and a large range of features collected per sample. In these cases, traditional methods of yield-prediction modeling become less robust. To handle this, a functional regression approach was employed that addresses limitations of previous techniques to create a model for predicting CMT, grain yield and other traits in wheat under heat stress environmental conditions and when data availability is constrained. The results preliminarily indicate that the overall models of each trait studied presented a good accuracy compared to their data's standard deviation. The yield prediction model presented an average error of 13.42%, showing the function-on-function algorithm chosen for the model as reliable for small datasets with high dimensionality.

Keywords Wheat · Cellular membrane thermostability · Grain yield · Function-on-function regression · Machine learning · UAV

Lucas Costa and Jordan McBreen have contributed equally.

✉ Yiannis Ampatzidis
i.ampatzidis@ufl.edu

✉ Md Ali Babar
mababar@ufl.edu

Extended author information available on the last page of the article

Introduction

Predicting yield early in wheat (*Triticum aestivum*) breeding programs can reduce the time it takes to produce a new variety and accelerate crop management practices over the season, like market planning, fertilizer application and harvest procedure preparation (Raun et al., 2001). Developing new and improved wheat varieties adapted to a wide range of environmental conditions in a timely manner is key to successful wheat breeding. However, accounting for the effects of the environment on the expression of a genotype as well as the high cost of multi-site evaluations in both time and labor spent can be a challenge. Statistics and genomic mechanisms ease these challenges through accurate predictions of grain yield and other parameters allowing for the rapid selection of candidate lines for target environments. Multiple approaches to grain yield prediction in agriculture have been developed, including those using soil analysis combined with plant indices (Raun et al., 2001; Pantazi et al., 2016), weather conditions and time-series analyses (Bell and Fischer, 1994), or multiple plant indices from canopy reflectance (Hansen et al., 2002). As these approaches rely on a small range of features that define the plant, they require large amounts of data to develop robust models that allow for the indirect selection of superior wheat lines through predictions of certain phenotypic parameters.

High-temperature stress in wheat can lead to shorter grain growth, causes early plant senescence and reduce photosynthesis, thus reducing the grain filling period, accelerating phasic development reducing carbon assimilation and reducing spike fertility, grain weight and yield (Farooq et al., 2011). These negative effects can be amplified when heat stress is experienced during the reproductive period (Midmore et al., 1982; Al-Khatib and Paulsen, 1984; Wardlaw et al., 1980). Thus, quantifying the heat stress tolerance (HST) of wheat lines is crucial for maximizing crop production. Heat stress tolerance can be thought of as a particular plant or genotype's capacity to maintain expected levels of grain yield and other quality traits while experiencing environmental temperatures considered to be above the threshold at which a wheat plant can be expected to experience high temperature physiological stress (Balla et al., 2011). The heat stress tolerance of a particular wheat line can be indirectly estimated by determining the relative performance of certain morphophysiological traits when exposed to temperatures higher than the typical range of growth. Various complex physiological and morphological traits are used for genetic improvement against heat stress, including cellular membrane thermostability (CMT), which is related to cellular thermo-tolerance and is estimated through the permeability of cellular membranes. CMT has been used as an indirect measurement of heat-stress tolerance and grain yield in diverse plant species, including wheat (Reynolds et al., 1998; Blum et al., 2001; Talukder et al., 2014).

In-field surveying for plant phenotypes and trait detection to assess plant yield status and create plant inventory are expensive, labor-intensive and time-consuming (Ampatzidis et al., 2017; Luvisi et al., 2016). High throughput phenotyping (HTP) shows potential to alleviate issues of phenotyping and improve wheat breeding efficiency (Babar et al., 2006; Babar et al., 2007; Rutkoski et al., 2016; Krause et al., 2019). Unmanned aerial vehicles (UAVs) equipped with sensing systems can collect and produce important phenotypic data much quicker than time-consuming in-situ phenotyping methods. Techniques for HTP utilizing UAV-collected spectral data and machine learning have been used for the rapid evaluation of several crops in a large-scale commercial field setting (Ampatzidis and Partel, 2019; Ampatzidis et al., 2019, 2020), including vegetation index development (Costa et al., 2020) and disease detection (Abdulridha et al., 2019, 2020a, b). In wheat, Zhang et al.

(2020) used multispectral satellite imagery to obtain a root mean squared error (RMSE) in predicting the yield of 786.50 kg/ha. Zhang et al. (2018) got an RMSE of 353.00 kg/ha on only the chosen spectra wavelengths. However, comparison between different studies is problematic since each has its own dataset. To precisely evaluate the best approach for predictions, the same dataset has to be used.

Hyperspectral UAV imagery involves complex and high-dimensional data. The unit of observation is a curve or set of finely sampled curves over a grid. Data analytics refer to these data points as functional data. It has been found that the appropriate way to process these functional data points is to integrate all of the information contained within collected spectral wavelengths into the prediction model to leverage latent data structures that otherwise would be lost if individual band-specific vegetation indices (VIs) are used instead (Aguate et al., 2017). When a large sample of high-dimensional data is available (i.e. there are a large number of individuals or plots from which spectral data are collected) that can be used as within both a training and testing dataset. Regularized neural networks have been shown to accurately model grain yield predictions in wheat (Ruß et al., 2008). However, in cases with fewer plots to collect data from and a narrower range of phenotypic values, these methods become less robust. To process such data, methods that can exploit the non-linear correlation structure of the spectral curves (aka. plant signatures) and consider the input-output correlation structure of data is desirable. Partial least square (PLS) is a well-established technique (Hansen et al., 2002) that captures the input-output correlation when performing dimensionality reduction. Nevertheless, PLS is a linear approach and does not exploit the non-linearity of the curves. Functional regression techniques are particularly designed for this purpose and can perform dimensionality reduction by exploiting non-linear correlations. However, most of the functional regression approaches used in yield prediction ignore the input-output relationship when performing the dimensionality reduction (Montesinos-López et al. 2017a, b). This paper employed a recently developed functional regression approach that addresses the limitations of the previous techniques to create a model for predicting grain yield in wheat under heat stress environmental conditions and when data availability is constrained by the processing of spectral data derived from the canopy reflectance of wheat plants collected via UAV-based hyperspectral sensing. With this in mind, the objective of this research was to successfully and tenably implement the novel functional regression approach in a heat-stressed winter wheat production system to achieve satisfactorily high prediction accuracies in spite of low data volumes.

Materials and methods

Plant materials and experimental design

Field evaluations of a US facultative soft wheat panel were carried out by characterizing several yield and heat stress tolerance contributing traits in the wheat plants. The panel consisted of 40 advanced genotypes selected from a larger population of a soft red facultative wheat association mapping panel (SWAMP) (Pradhan et al., 2019) known to be adapted to the warm and humid south and southeastern regions of the USA. The vernalization requirements for those genotypes are generally low. The experiment was grown at Plant Science Research and Education Unit, or PSREU, in Citra, Florida, during the 2018–2019 season. The wheat plots in Citra (29°24′17.6″N 82°10′41.5″W) were planted in a 5.09 m² plot (3.35 m long/1.52 m wide) using an α -lattice experimental design (2

replications per planting), resulting in a total of 160 plots. The wheat plots were machine planted and consisted of seven equally spaced rows. Machine harvesting was carried out when plots had reached maturity, and grain moisture level was at around 12%. Harvesting dates were May 9th and May 14th, 2019, for the early and late planting dates. The panel was genetically characterized using 27,244 GBS-SNPs markers.

During the post-anthesis/grain-filling stages of wheat development, the temperatures in Citra, Florida fall between 30 and 35 °C, typically higher than most of the other wheat-growing areas of the United States. The experiment was planted on November 15th, 2019, for early planting, and on December 15th, 2019, for late planting conditions. This way, the panel planted under the late planting condition could maximize the heat stress experienced during the reproductive and grain filling growth stages for those plants. The wheat plots in this experiment were irrigated as needed to avoid any unwanted water stress through overhead center-pivot irrigation every 15 days if there was no occurrence of rain. For fertilization, nitrogen (N), phosphorus (P) and potassium (K) were applied at the rate of 5-10-15 plus sulphur and micronutrients on the day of planting. Additionally, N was applied as a top dress as needed at a rate of 100.00 kg/ha. Round-up® (Bayer CropScience, Leverkusen, Germany) was applied at the rate of 1.120 kg/ha at the fourth week after planting. Harmony® Extra (Dow Chemical Company, Midland, Michigan, US) was applied at the rate of 42.03 g/ha six weeks after planting. Machine or manual weeding was carried out when required. Prosaro® 421 SC fungicide (Bayer CropScience, Leverkusen, Germany) was applied at the rate of 560.43 g/ha to control foliar diseases.

Trait measurement

A number of agronomic and heat stress tolerance secondary indicator traits were measured during the growing season, including grain yield measured in kg/ha, thousand kernel weight or TKW measured in grams, grain number per meter squared or Gnm^{-2} and cellular membrane thermostability or CMT, which is a unitless measurement by its nature. Grain number per meter squared was included over other traits as it represents the combination of spike number and grains per spike traits and thus is the most important yield component. This trait is affected by high temperature stress causing floret abortion during the flowering stages of the wheat plants, thus justifying the inclusion of Gnm^{-2} in the study. It is worth noting that certain traits in wheat such as spike number are important yield component traits, but are typically more affected by the environment as well as the ratio between fertile and non-fertile tillers. CMT has been shown to be a good heat stress tolerance indicator trait as it illustrates a plant's ability to resist the protein denaturing that occurs as a result of heat shock damage to the cellular membrane. CMT is best understood as the difference in electrolyte leakage of leaf tissue samples before and after a heat shock autoclaving event. To determine a wheat genotype's CMT, ten 10 mm leaf discs were collected from flag leaves of 10 individual plants chosen at random per experimental plot using a normal paper hole-punch. The 10 leaf discs were collected midway between the base and the tip of the leaf blade and placed into a glass vial containing 20 ml of ice-cold deionized water. The vials were kept on ice in a cooler leaf disc perturbation and electrolyte leakage is minimized. Once the vials containing the samples were back in the lab, they were placed on a shaker at room temperature for 24 h to stimulate electrolyte leakage. After the first 24 h on the shaker, the electrolyte leakage readings were taken using a benchtop conductivity meter (Orion Star™ A212 Conductivity Benchtop Meter, Thermo Scientific™, Waltham, MA, USA). Immediately after, samples are autoclaved to kill the leaf tissue and release any

remaining electrolytes. They are then put back on the shaker for another 24 h at room temperature. After the second 24 h, the electrolyte leakage readings were retaken. CMT will then be expressed in percentage units as the reciprocal of relative electrolyte leakage. CMT is then calculated from the two readings as:

$$\text{CMT} = (1 - \text{Initial Reading} / \text{Second Reading}) * 100 \quad (1)$$

CMT was taken at two different time points: 7 days after anthesis and 14 days after anthesis. The average of these two values for each plot was then used as the overall CMT.

Yield in kg/ha and yield-related traits, including grain number per square meter (Gnm^{-2}) and thousand kernel weight in grams (TKW) were also calculated at maturity. Yield data was taken from the machine-harvested grain yield and was adjusted to a 12% moisture level and was used to calculate Gnm^{-2} . For TKW, 250 kernels were collected from 10 wheat spikes randomly selected per experimental plot. The grain weight for 250 wheat seeds or kernels was determined and converted into 1 000-grain weight. Thus, 1 000 kernel weight (TKW) is calculated as:

$$\text{TKW} = (\text{Grain Weight} / \text{Seed Count}) * 1000 \quad (2)$$

Gnm^{-2} was determined for each individual by the following equation:

$$\text{Gnm}^{-2} = (\text{Grain Yield} / 10) / (\text{TKW} / 1000) \quad (3)$$

Heading, anthesis and maturity date data were also taken as the number of days from the initial planting at which 50% of the plants within a plot had headed, flowered or yellowed/senesced, respectively.

UAV-based hyperspectral imaging

Spectral data were collected using a hexacopter UAV (Matrice 600 Pro, DJI, Shenzhen, China) equipped with a hyperspectral sensing system, Pika L 2.4 (Resonon, Bozeman, MT, USA). For UAV flight planning and mission control, the Pix4DCapture (Pix4D S.A., Prilly, Switzerland) software app was used on an iPad (Apple, Cupertino, CA, USA) connected to the remote controller of the UAV. The hyperspectral sensing system covers a 400–1 000 nm range divided into 150 bands. It uses a 17 mm focal length lens with a field of view (FOV) of 17.60° and an instantaneous field of view (IFOV) of 0.71 mrad. In addition, this sensing system includes an inertial measurement unit (IMU) (DJI, Shenzhen, China) for multi-rotor aircraft flight control, recording the sensor position and orientation. The hyperspectral system has its own software and works separately from Pix4D. A front overlap of 85% and side overlap of 70% were used for spectral data collection. Data were collected at an altitude of 60.96 m above the ground. A white calibration tarp was placed in the data collection region to calibrate the hyperspectral data collected.

The hyperspectral data were calibrated and analyzed using the Spectronon software (Spectronon Pro, Resonon, Bozeman, MT, USA). The georectification and radiometric corrections were performed using the Spectronon Pro software, to get the GNSS/IMU and radiometric data. The regions of interest (RoI) were manually selected based on each plot position, and the spectral wavelengths data were collected. Hyperspectral imagery was collected post-anthesis during the grain filling stages, and data taken from the best flight were used to develop the model for trait prediction. Best spectral data corresponded with the soft dough wheat growth stage, at or around Zadoks 85 (Simmons et al., 1985).

Function-on-function regression

Functional (profile) data refers to a set of observations that are temporally or spatially correlated. Figure 1 illustrates an example of functional data with the spectral reflectance curve of a winter wheat sample. Spectral data measures the light being reflected by objects, with reflectance being the proportion of light radiation striking a surface reflected off it, given in percentage. Each pixel of a hyperspectral image is associated with a profile where the value of reflectance is observed at each wavelength. The reflectance values are correlated, meaning that the neighboring wavelengths show close values of reflectance. Because of this correlation structure, treating each reflectance value as a separate variable (as most machine learning algorithms, including linear regression, do) causes overfitting and low predictability power. Therefore, algorithms that perform dimensionality reduction and summarize the profile data into a few features are required. Principal component analysis (PCA) as a method of dimensionality reduction of spectral profiles could be employed. However, it is not clear whether the features obtained from the dimensionality reduction step are strong predictors of the yield, as the measured errors correspond to the dimensionality reduction and regression steps together. As a result, more advanced approaches that perform dimensionality reduction simultaneously with regression modeling are necessary. Partial least squares address this limitation but fail to capture the non-linear correlation structure in profile data. Functional regression approaches have been designed to address this limitation. These models can include all the observations within a profile as follows:

$$y_i = \alpha + \int \beta(t)x_i(t)dt + e_i, \quad (4)$$

where α and $\beta(t)$ are model parameters to be estimated; $x_i(t)$ is the value of the profile at the t^{th} spectral frequency for the i^{th} sample; and y_i is the yield value to be estimated.

Fig. 1 Example of reflectance signature in winter wheat

Most of the existing functional regression approaches use a predefined set of bases, including Fourier and splines, to span the functional space and decompose the curves. However, the selection of these bases is arbitrary, and it is not clear which set of bases are most appropriate. Therefore, data-driven learning of these bases while capturing the input-output relationships is desirable. A recently developed approach is called function-on-function regression with signal compression (Sigcomp), which identifies these bases optimally (Luo and Qi, 2017). It is superior to most existing functional regression approaches in terms of prediction accuracy. Sigcomp was employed in this study for yield prediction purposes.

Cross-validation clusters

To validate the regression model, a 5-fold cross-validation scheme was considered. To control the genomic relatedness among lines, the population was firstly stratified into 10 clusters based on a discriminant analysis of principal components (DAPC) (Jombart et al. 2010) analysis using all single-nucleotide polymorphism (SNP) markers. Approximately 4–6 lines were randomly drawn from each cluster to compose one of five new sub-groups for cross-validation. This process was replicated 10 times for model training. Figure 2 shows the representation of the division of the dataset into five sub-groups.

Experimental design

Data analysis

The dataset was analyzed using classical statistical techniques to obtain the mean, standard deviation, skewness and coefficient of variation (CV). The standard deviation measures the spread of a distribution compared to the mean value, where lower (close to zero) values represent that the data points are closer to the mean. The skewness evaluates the data distribution, where values close to zero correspond to a symmetric distribution. Skewness presents a distribution of values clustered on smaller values (negative skewness) or higher

Fig. 2 Representation of the division of the dataset into five groups based on 10 clusters according to similar genomic data

values (positive skewness). The coefficient of variation shows the variability of the dataset to its mean value in percentages.

Yield traits prediction model

In this study, spectral data were collected by the UAV-based hyperspectral sensing system from the wheat canopy. Manual data were also collected, such as days to heading and grain yield per area for each wheat plot. A model was created for each trait of the dataset. The dataset was split into five groups based on a previous clusterization of the genotypes by genomic data to allow each group to have genetically similar populations. A predictive model was trained by the Sigcomp algorithm using four of the groups as a training dataset and test the model's performance on the remaining group. This procedure is repeated five times as is called cross-validation. This model predicts yield values based on the spectral data collected on April 23rd, 2019, approximately one month before harvest. Ideally, data from multiple flights with the hyperspectral camera-equipped UAV would be used for data processing. However, due to UAV availability and sensor miscalibration issues, only the data from a single flight date were usable for both the early and late plantings. The values for the predicted trait were then used together with the ground truth to measure the accuracy, repeatability and reproducibility of the created model.

Evaluation metrics

Evaluation metrics are used to compare the proposed solution with other works and the current methodology. Errors are used to generate the expected precision of a solution when used in a group of individuals, as they converge to the calculated error with larger numbers of measurements. The average error is useful when estimating values by averaging samples.

The measure of agreement is a comparison between two methodologies and generates an operational error within a confidence interval. The operational error is defined by a maximum and minimum error for which any subsequent estimation may lie. This means that there is an X percent chance for any subsequent usage of such methodology (where X is the confidence interval) that the measurement error will be within the maximum and minimum values.

Mean absolute percentage error (MAPE) and root mean squared error (RMSE)

The mean absolute percentage error (MAPE) is the computed average of absolute percentage errors by which forecasts of a model differ from actual values of what is being forecast. In this study, the yield prediction from the model is compared to the actual ground truth of yield gathered from harvest. The MAPE is given by the absolute difference between both the ground truth (Gt) and the prediction (P), divided by the ground truth, divided by the number of individuals (n). The formula for the MAPE is presented in Eq. 5.

$$MAPE = 1/n \sum |(Gt - P)/Gt| \quad (5)$$

The root mean squared error (RMSE) is the root of the squared difference between the ground truth (Gt) and predicted value (P), divided by the number of individuals (n) (Eq. 6). This is heavily used in the literature to measure the accuracy of a prediction model.

$$RMSE = \sqrt{((Gt - Pv)/n)} \quad (6)$$

Measure of agreement

The agreement between measurements refers to the degree of concordance between two sets of measurements of the same individual by two different methodologies. Statistical methods to test agreement are used to decide whether one technique for measuring a variable can substitute another (Ranganathan et al., 2017). In this study, the methodology imposed is the estimation of yield from the spectral data of the wheat canopy compared to the actual values from harvest. In such a case for prediction, the value of reproducibility and repeatability is also important, as it gives a range of operation to which the model can be applied.

The Pearson correlation coefficient is often inappropriately used to evaluate for agreement, since it is an incorrect measure of reproducibility or repeatability (Watson and Petrie, 2010). Observations of two different methods are often compared with this method, which only gives a percentage (r^2 in this case) that a new value would be inserted in the trending line, but does not give a range of certainty on which it would be predicted, or how accurate the real prediction is actually made.

The Bland and Altman diagram (Bland and Altman, 1999) is a display of the pattern and agreement of one variable being measured by two different methodologies (Watson and Petrie, 2010). The diagram plots the difference between a measurement pair on the vertical axis and the mean of the pair on the horizontal axis. To determine the repeatability, the method assumes a normal distribution of differences, where 95% of them are expected to lie between $d \pm 1.96 s$, where d is the mean of observed differences and s is the standard deviation. The value of 1.96 represents 95% of the area of a normal distribution curve, where $d \pm 1.96 s$ represents the lower and upper bounds of the range. The Bland and Altman diagram has a limitation of use in cases where a funnel effect is observed, where the data in the diagram resembles a funnel because, on higher mean values, the difference is lower. In such cases, the dataset must be appropriately transformed before use.

Results

Dataset analysis

The results of the dataset analysis are presented in Table 1. The values of coefficient of variation shows a variability of around 15% for most of the traits. The skewness also shows a distribution with low concentration of values in a certain range, as most values are close

Table 1 Dataset analysis for each of the traits

Trait	Min–Max	Mean	CV (%)	Skewness
Yield [kg/ha]	2413–7675	5279	15.79	−0.51
CMT	50.78–92.18	74.44	7.30	−0.97
Gnm ² [gn/m2]	8917–26,310	17918.22	17.82	0.02
TKW [g]	19.60–40.40	29.64	13.82	0.31

to zero. The CMT trait presents lower variability (lower CV) and higher concentration of values in a certain range (Skewness close to -1). This makes the CMT trait model more susceptible to overfitting.

The dataset analysis is important to show the context for each model created. As most traits are expected to have similar values from a study with the same conditions, some variance is still expected to create a good model and avoid overfitting. These results suggest the need for a closer evaluation of the CMT model created.

Cross-validation grouping analysis

The dataset was divided into similar groups based on genomic information. Therefore, it is expected that the average spectral curve representing each group to be similar. Figure 3a shows the average spectral curve for each group. This shows that groups 4 and 5 have differences in the spectrum, which can be traced to differences in the sunlight condition during the flights. To counter the weather conditions on multiple flights, the curve is normalized by dividing every wavelength by the reflectance value of the first. The normalized curves are presented in Fig. 3b. The normalized curves present the similarity expected between the groups, with some small variation on group 5 that is still acceptable.

Like the spectral curves, the average value for all traits in each group is also expected to be similar. To demonstrate the similarity of each trait, they are normalized by dividing each one by the maximum value. Figure 4 presents the normalized average trait values for each group. This shows that the average traits of each group vary under 10% of the difference. From both Figs. 3 and 4 analysis, it can be concluded that the population of each group is similar and therefore ideal for cross-validation.

Yield prediction

The panel consists of 144 plots after eliminating plots with no yield measurements collected. On all five folds of the cross-validation, the MAPE for each group is presented in Table 2. Overall, the archived MAPE achieved was 13.42%. Such a low error indicates that given the relatively low number of individuals, the model is still robust for new data, with good repeatability as all the MAPEs found are relatively close.

The Bland and Altman diagram for the panel is presented in Fig. 5. It shows a reliability range of -1663.97 to $+1754.56$ kg/ha range of error. This means that the model can predict the yield within around 32% error (using the mean value as a base) on 95% of the predictions. The range of 32% in error with high reliability demonstrates the potential of this model for even datasets with a low number of individuals.

CMT prediction

The CMT model MAPEs are presented in Table 3. The average MAPE for all cross-validation folds presents a value of 5.28%. The low MAPE is predicted from the low variability of the dataset. To evaluate for an overfitting case, the variance of the dataset can be used. The average MAPE of 5.28% is equivalent to a variation of 3.93 of CMT units, which is lower than the variance of 10.19 of the dataset. This means that the result achieved is more precise than the variation of the dataset, demonstrating that the model has high prediction power and is not overly fitted.

Fig. 3 **a** Mean spectral curve for each group, **b** normalized mean spectral curve for each group

The Bland and Altman diagram for the CMT model is presented in Fig. 6. This shows a range of operation errors of -10.13 to $+10.21$. This range of operation corresponds to around 13.71% of error considering the mean value of the CMT data. However, this range of operation is still precise, with under 15% error for all expected measurements using spectral data.

Fig. 4 Normalized average trait values for each group on the 5-fold cross-validation

Table 2 Value for MAPE, RMSE, standard deviation, minimum and maximum error values for each group in the cross-validation for yield prediction

	Group 1	Group 2	Group 3	Group 4	Group 5	Average
MAPE	15.61%	10.18%	12.20%	12.03%	17.10%	13.42%
RMSE	960.44 kg/ha	757.06 kg/ha	845.39 kg/ha	792.02 kg/ha	991.08 kg/ha	869.20 kg/ha
Standard Deviation	10.83%	8.96%	12.82%	6.73%	10.40%	9.95%
Min - Max	0.72–40.71%	1.02–35.86%	1.07–56.12%	0.56–25.15%	0.26–40.13%

Fig. 5 Bland and Altman diagram for the yield prediction model

Table 3 Value for MAPE for each group in the cross-validation for CMT estimation

	Group 1	Group 2	Group 3	Group 4	Group 5	Average
MAPE	5.05%	4.09%	5.52%	4.24%	7.47%	5.28%
RMSE	5.03	4.14	4.76	4.08	7.23	5.05
Standard Deviation	3.73%	3.88%	3.92%	3.69%	3.97%	3.84%
Min–Max	0.04–16.82%	0.02–16.01%	0.75–14.97%	0.14–16.05%	0.09–15.42%

Fig. 6 Bland and Altman diagram for the CMT dataset**Table 4** Value for MAPE for each group in the cross-validation for Gnm^2 estimation

	Group 1	Group 2	Group 3	Group 4	Group 5	Average
MAPE	17.80%	11.68%	13.16%	12.61%	18.66%	14.73%
RMSE	3,541.79	2,548.85	3,063.61	3,166.15	3,597.14	3,183.51
Standard Deviation	13.80%	7.99%	12.07%	12.17%	10.91%	11.39%
Min–Max	0.00–49.45%	0.94–41.02%	0.99–53.58%	0.34–60.39%	1.77–56.13%	

Gnm^{-2} prediction

The MAPEs for the Gnm^{-2} model are presented in Table 4. The average MAPE for all cross-validation folds is 14.73%. This represents a good model for such a small training dataset. The Gnm^{-2} trait has the largest standard deviation of all data but still has a low error similar to the other traits. This shows the high accuracy of the algorithm chosen for modeling.

It can be seen that with the larger CV of the data, the model is expected to have a larger spread of errors. However, the MAPE is still lower than 20% on all folds. The

Fig. 7 Bland and Altman diagram for the Gnm^{-2} model

Table 5 Value for MAPE for each group in the cross-validation for TKW estimation

	Group 1	Group 2	Group 3	Group 4	Group 5	Average
MAPE	8.69%	7.60%	8.93%	7.95%	12.39%	9.13%
RMSE	3.28 g	2.88 g	3.37 g	2.80 g	4.22 g	3.31 g
Standard Deviation	6.55%	5.47%	5.85%	5.68%	10.01%	6.71%
Min–Max	1.49–30.06%	0.09–22.99%	0.72–24.70%	0.05–19.86%	0.17–37.25%	

low MAPE with a large spread of errors shows that the model algorithm is reliable on all sizes of datasets. This shows a promising application for further studies.

The Bland and Altman diagram for the Gnm^{-2} model is presented in Fig. 7. It shows a reliability of error between -6125.15 and $+6414.87$. When comparing the mean value of the data, these values correspond to an error of operation of 35.82%. This is a good estimation for a model trained in a small dataset.

TKW prediction

The cross-validation MAPE for the TKW model is presented in Table 5. The overall average value of MAPE for this model is 9.13%. This low MAPE compared with the CV of the dataset presents a good accuracy for the model in predicting this trait.

The Bland and Altman diagram for the TKW prediction model is presented in Fig. 8. The diagram presented a margin of error of operation of -6.56 to $+6.59$ units of TKW. Using the mean value of the trait, this range is equal to a range of -22.13% to $+22.23\%$. Thus, this error of operation is still effective.

Fig. 8 Bland and Altman diagram for the TKW prediction model

Discussion

Increasing global temperatures as a result of global climatic disruptions, especially in subtropical and semi-arid latitudes, has been projected to have detrimentally adverse repercussions on the productivity of wheat breeding and production; with climatic modeling studies predicting more severe and frequent heatwaves in the near future (Ortiz et al., 2008; Trnka et al., 2014). High temperatures can hinder the development and growth of wheat plants during both the day and nighttime at all developmental stages, but the adverse effects of high-temperature stress during the grain filling stage, in particular, has been shown to have deleterious effects on growth and ultimately grain yield (Reynolds et al., 2012). Yield, yield-related traits and complex physiological trait (CMT), an indicator of heat stress tolerance in wheat, were explored to develop a model for wheat yield prediction under high-temperature environmental conditions and relatively small-scale data availability.

The ground-collected phenotypic data and hyperspectral imaging data were derived from a 40-genotype soft red facultative wheat panel. The genotypes have been well-characterized as being adapted to the high temperatures and humidity experienced by south and southeastern states in the USA. The 40-genotypes chosen for this experiment were specifically selected to encompass most of the genetic variability within the south and the southeastern USA. The reasoning behind only using a small, 40-genotype subset of a greater genotypic population was to simulate grain breeding scenarios where very large-scale experiments with abundant genetic material are infeasible due to financial, environmental or phenotypic data taking constraints. Such is the case for many breeding programs in still-developing areas of the world that share subtropical, high-temperature climates. Multi-family populations have been proven to be optimal for developing prediction models in small grain crops compared to using bi-parental populations as they can facilitate the incorporation of greater genetic variability into a model. However, a major drawback to using multi-family populations is that they tend to require large population sizes, extensive environmental data and dense marker genotypes. With this goal and apparent limitation of small population size in mind, the function on function linear regression with signal compression approach was taken on a multi-family population. This study demonstrates that UAV-based

hyperspectral imaging is a promising high throughput phenotyping tool for predicting genotypic performance for grain yield and complex physiological traits in wheat in a small set of genotypes under stress conditions. The combination of hyperspectral phenotype (which is associated with complex in-season physiological and yield-related traits) with genomic information has the potential to improve the accuracy and quality of phenotypic prediction and will enable breeders to capture the genetic potential of wheat to cope with heat stress.

Overall, the models of each trait presented a good accuracy compared to their data standard deviation. This shows that the function-on-function algorithm chosen for the model is reliable for small datasets with large amounts of data. The MAPE is a good tool to evaluate the accuracy of the model on a large-scale experiment, but the Bland and Altman diagram shows a better evaluation of the error using metrics of standard deviation, on small-scale experiments. This presents a better range of error when the model is in actual operation. Unfortunately, only the spectral data from just one of the multiple flight dates that were conducted was used due to issues with the hyperspectral scanner itself when conducting aerial data collection and limited availability to conduct the hyperspectral flights. In an ideal scenario, the data from the multiple flights that were conducted would be used. Despite this, the findings act as an inaugural study into demonstrating the effectiveness of the application of a function on function linear regression with Sigcomp to high throughput phenotyping in wheat breeding. Future studies will involve incorporating spectral data from multiple flight dates from multiple growth stages of the wheat plants to ascertain the relationships between phenotype and spectral reflectance over the plant's life cycle for early grain yield prediction.

To determine model prediction accuracy, the 40-genotype panel used in this study was partitioned into five subgroups of equivalent size, where four subgroups are used to fit the prediction model, serving as the training population. The other leftover subgroup served as the training or validation population to assess any correlations among the observed and predicted values of the traits being studied. The process is repeated five times so that each of the five different subgroups can be used as the prediction set at least once. In order to control the genotypic relatedness among lines, the panel was stratified based on the discriminant analysis of principal components (DAPC) clustering analysis to ensure that lines included within the same group are present in either the training population or validation population, but never in both concurrently.

Of the traits examined in this study, the heat stress tolerance indicator physiological trait CMT presents a lower variability than the other traits examined and a higher concentration of values within a certain range, leading to a skewness value very near -1.00 (-0.97). Because of this skewed nature of the CMT data, the developed model for the trait is more susceptible to overfitting issues. This may indeed indicate that there is a need for a closer evaluation of the CMT model that was created in this experiment. This issue regarding overfitting for CMT may be ameliorated by incorporating CMT data points from less or similarly related wheat genotypes that exhibit more extreme values for CMT, that is to say, lines with very high and very low CMT values. Multi-year and multi-environment data for CMT may also help to solve issues of model overfitting for the trait. Despite these prospective issues that were foreseen to complicate predictions of CMT, the model still achieved a highly precise range of operation (between -10.13 and $+10.21$) with under 15% error for all expected measurements utilizing spectral data, as shown in the Bland-Altman diagram of Fig. 5. Also, for CMT, the average MAPE for all cross-validation folds was 5.28%, which is very low but was so due to the low variability exhibited by the dataset and did not speak to the robustness of the model. Because of such low variability in CMT values within the dataset, to avoid overfitting, it was instead opted to use the variance of the CMT

dataset; where the average MAPE of 5.28% corresponds to a variation of approximately 3.93 CMT units, much lower than the variance of the dataset itself which was 10.19 CMT units. This would indicate that the model developed has relatively high prediction power without overfitting the data. However, because of the relative difficulty and labor intensity of collecting CMT measurements, studies incorporating this secondary indicator trait are scarce. Thus, the results from this study offer novel insight into the potential of this trait to be further incorporated into grain yield estimations in wheat crops.

For grain yield and the yield-related traits of Gnm^{-2} and TKW, the average MAPEs for the models were consistently under 15% error (13.42%, 14.73 and 9.13%, respectively). This low level of error for all three traits indicates that despite only a relatively small number of individuals present within the dataset, the model that employs function on function linear regression with Sigcomp is still robust for incorporating new data. The MAPEs for each trait are also relatively close to one another, indicating that the models exhibit good repeatability. For grain yield, a reliability range of -1.66 to $+1.75$ t/ha in the Bland-Altman diagram indicates the model's capacity to predict yield with $\sim 32\%$ error on 95% of the predictions when using the mean value as a base. A range of 32% in error with high reliability is acceptable for yield and demonstrates that this model could potentially be applied to datasets with even fewer individuals. In comparison to other works in the literature for yield prediction, the current approach got a RMSE of 869.20 kg/ha. Zhang et al. (2018) got an RMSE of 353.00 kg/ha, but used a spectroradiometer, which collects spectral data at the leaf level and requires a ground-based measurement. Zhang et al. (2020) got an RMSE of 786.50 kg/ha by using satellite spectral imagery and used only a five-point sampling to represent the whole area. Similar studies that assimilated crop phenology and LAI into crop growth models for grain yield estimation in wheat achieved RMSE values of 737.00 kg/ha (Chen et al., 2018) and 585.00 kg/ha (Huang et al., 2015) when using remote sensing data derived from satellite imaging. These error values are lower than those achieved in this study but are still comparable when considering the large number of data points used in those studies compared to the relatively few available in the present study. Gaso et al. (2019) achieved RMSE for estimated yield in wheat of 966.00 kg/ha when applying a simple regression estimation method based on vegetation indices from satellite imaging data, suggesting that UAV-based imaging and yield estimation approaches as potentially superior. In general, comparison between different studies based on RMSE is problematic since each study has a different dataset.

In the case of Gnm^{-2} , the data presented a larger spread compared to any of the other investigated traits. However, it still had a low error (14.73%) similar to the other traits. This low MAPE with a large spread of errors proves the model to be promising for application in future studies as it is reliable on different sizes of datasets. The reliability of error for the Gnm^{-2} model was between -6125.15 and $+6414.87$, which corresponds to an error of operation of 35.80%, which is acceptable for a model trained on a small sized model dataset. TKW had the lowest average MAPE of 9.13%, which, when compared with the CV of the dataset, presents a good accuracy for the model's prediction of the trait. When looking at the Bland-Altman diagram for TKW, the margin of error of operation is -6.56 to $+6.59$ units of TKW. When using the mean value of the trait, this range is equal to -22.13% to $+22.23\%$, which is still effective for prediction purposes.

This particular research experiment fits into the greater body of high throughput phenotyping for agronomic crop research. It provides a snapshot of the capabilities of hyperspectral UAV-based imaging when applied to phenotyping to make connections between phenotypic traits and the genotypes that drive their expression. Standardization of small-grain agronomic crop phenotyping has proven a unique challenge due to the

many differences between target environments and interaction effects that are bound to occur between the genotype and phenotype and the genotype and the growing environment. Although there are studies that tackle this issue, such as Guo et al. (2020) and Lado et al. (2016), it is emphasized in limited data scenarios such as the one presented in this research. The methods and model presented in this study have the potential to be generalized to a larger dataset containing more diverse wheat genotypes that are grown in multiple environments, which would support the concept of transforming novel crop HTP collection and analysis methods into standardizable data processing pipelines that can break into the world of commercial crop breeding. Over the years, aerial-based high throughput crop phenotyping from a data collection perspective has adopted specific techniques as standard approaches, such as the usage of crop vegetation indices as metrics of plant health, the employment of reflectance panels in the field for radiometric image correction and the implementation of ground control points for accurate image georeferencing. However, the implementation and adoption of different models for phenotypic trait prediction in wheat remain largely varied; thus accentuating the importance of the current study in making steps towards increasing the generalizability of high throughput phenotyping crop prediction models.

Conclusions

Precise and reliable models for yield traits prediction could help researchers find stress tolerant germplasm and link genotypic variations to natural variations and adaptive traits under field conditions. Yield traits prediction models for wheat are developed and evaluated as to their accuracy and reliability. This study illustrates the use of hyperspectral UAV systems to predict yield and certain other traits (grain number per square meter, thousand kernel weight and cellular membrane thermostability) for facultative wheat genotypes under conditions where data is constrained and data points are limited. Here, the values for yield and three other traits were processed by a function on function linear regression with signal compression to estimate the precision of yield and other trait predictions through UAV-based hyperspectral imaging techniques. The results presented from the data proved to be promising, giving relatively high accuracies for all the traits considered compared to each of their data standard deviations. Yield and other trait prediction through the function on function algorithm employed here denote a marked advancement in the ability to develop models with high prediction accuracies when relatively little phenotypic data is available and when that data does not necessarily encompass a wide range of values. Predicting wheat yield and performance under heat stress through advanced statistical predictive models allows breeding programs to make line selections earlier than normal in the breeding cycle, therefore increasing the speed of cultivar development and release; now it has been shown that this is possible even when the breeder does not have copious amounts of phenotypic data. Future research can be oriented towards incorporating phenotypic data of other secondary traits related to high yield under heat-stressed conditions into the function on function model. Comprehensive studies on how UAV-based spectral data taken from early-stage wheat plants can be modeled to predict yield and other traits are also important to increase the selection speed in wheat breeding programs.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abdulridha, J., Ampatzidis, Y., Kakarla, S. C., & Roberts, P. (2020a). Detection of target spot and bacterial spot diseases in tomato using UAV-based and benchtop-based hyperspectral imaging techniques. *Precision Agriculture*, 21(5), 955–978. <https://doi.org/10.1007/s11119-019-09703-4>
- Abdulridha, J., Ampatzidis, Y., Roberts, P., & Kakarla, S. C. (2020b). Detecting powdery mildew disease in squash at different stages using UAV-based hyperspectral imaging and artificial intelligence. *Biosystems Engineering*, 197, 135–148. <https://doi.org/10.1016/j.biosystemseng.2020.07.001>
- Abdulridha, J., Batuman, O., & Ampatzidis, Y. (2019). UAV-based remote sensing technique to detect citrus canker disease utilizing hyperspectral imaging and machine learning. *Remote Sensing*, 11(11), 1373. <https://doi.org/10.3390/rs11111373>
- Aguate, F., Trachsel, S., Pérez, L., Burgueño, J., Crossa, J., Balzarini, M., et al. (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>
- Al-Khatib, K., & Paulsen, G. (1984). Mode of high temperature injury to wheat during grain development. *Physiologia Plantarum*, 61(3), 363–368.
- Ampatzidis, Y., De Bellis, L., & Luvisi, A. (2017). iPathology: robotic applications and management of plants and plant diseases. *Sustainability*, 9(6), 1010.
- Ampatzidis, Y., & Partel, V. (2019). UAV-based high throughput phenotyping in citrus utilizing multispectral imaging and artificial intelligence. *Remote Sensing*. <https://doi.org/10.3390/rs11040410>.
- Ampatzidis, Y., Partel, V., & Costa, L. (2020). Agroview: Cloud-based application to process, analyze and visualize UAV-collected data for precision agriculture applications utilizing artificial intelligence. *Computers and Electronics in Agriculture*. <https://doi.org/10.1016/j.compag.2020.105457>.
- Ampatzidis, Y., Partel, V., Meyering, B., & Albrecht, U. (2019). Citrus rootstock evaluation utilizing UAV-based remote sensing and artificial intelligence. *Computers and Electronics in Agriculture*. <https://doi.org/10.1016/j.compag.2019.104900>.
- Babar, M., Reynolds, M., Van Ginkel, M., Klatt, A., Raun, W., & Stone, M. (2006). Spectral reflectance indices as a potential indirect selection criteria for wheat yield under irrigation. *Crop Science*, 46(2), 578–588. <https://doi.org/10.2135/cropsci2005.0059>
- Babar, M., Van Ginkel, M., Reynolds, M., Prasad, B., & Klatt, A. (2007). Heritability, correlated response, and indirect selection involving spectral reflectance indices and grain yield in wheat. *Australian Journal of Agricultural Research*, 58(5), 432–442. <https://doi.org/10.1071/ar06270>
- Balla, K., Rakszegi, M., Li, Z., Bekes, F., Bencze, S., & Veisz, O. (2011). Quality of winter wheat in relation to heat and drought shock after anthesis. *Czech Journal of Food Sciences*, 29(2), 117–128. <https://doi.org/10.17221/227/2010-CJFS>
- Bell, M., & Fischer, R. (1994). Using yield prediction models to assess yield gains: A case study for wheat. *Field Crops Research*, 36(2), 161–166. [https://doi.org/10.1016/0378-4290\(94\)90064-7](https://doi.org/10.1016/0378-4290(94)90064-7)
- Bland, J., & Altman, D. (1999). Measuring agreement in method comparison studies. *Statistical methods in medical research*, 8(2), 135–160. <https://doi.org/10.1177/096228029900800204>
- Blum, A., Klueva, N., & Nguyen, H. T. (2001). Wheat cellular thermotolerance is related to yield under heat stress. *Euphytica*, 117(2), 117–123. <https://doi.org/10.1111/j.1439-037x.2009.00398.x>
- Chen, Y., Zhang, Z., & Tao, F. (2018). Improving regional winter wheat yield estimation through assimilation of phenology and leaf area index from remote sensing data. *European Journal of Agronomy*, 101, 163–173. <https://doi.org/10.1016/j.eja.2018.09.006>
- Costa, L., Nunes, L., & Ampatzidis, Y. (2020). A new visible band index (vNDVI) for estimating NDVI values on RGB images utilizing genetic algorithms. *Computers and Electronics in Agriculture*. <https://doi.org/10.1016/j.compag.2020.105334>.

- Farooq, M., Bramley, H., Palta, J. A., & Siddique, K. H. (2011). Heat stress in wheat during reproductive and grain-filling phases. *Critical Reviews in Plant Sciences*, 30(6), 491–507. <https://doi.org/10.1080/07352689.2011.615687>
- Gaso, D., Berger, A., & Ciganda, V. (2019). Predicting wheat grain yield and spatial variability at field scale using a simple regression or a crop model in conjunction with Landsat images. *Computers and Electronics in Agriculture*, 159, 75–83. <https://doi.org/10.1016/j.compag.2019.02.026>
- Guo, J., Pradhan, S., Shahi, D., Khan, J., Mcbreen, J., Bai, G., et al. (2020). Increased prediction accuracy using combined genomic information and physiological traits in a soft wheat panel evaluated in multi-environments. *Scientific Reports*, 10(1), 1–12. <https://doi.org/10.1038/s41598-020-63919-3>
- Hansen, P., Jørgensen, J., & Thomsen, A. (2002). Predicting grain yield and protein content in winter wheat and spring barley using repeated canopy reflectance measurements and partial least squares regression. *The Journal of Agricultural Science*, 139(3), 307–318. <https://doi.org/10.1017/s0021859602002320>
- Huang, J., Tian, L., Liang, S., Ma, H., Becker-Reshef, I., Huang, Y., et al. (2015). Improving winter wheat yield estimation by assimilation of the leaf area index from Landsat TM and MODIS data into the WOFOST model. *Agricultural and Forest Meteorology*, 204, 106–121. <https://doi.org/10.1016/j.agrfor.2015.02.001>
- Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC genetics*, 11(1), 94. <https://doi.org/10.1186/1471-2156-11-94>
- Krause, M., González-Pérez, L., Crossa, J., Pérez-Rodríguez, P., Montesinos-López, O., Singh, R., et al. (2019). Hyperspectral reflectance-derived relationship matrices for genomic prediction of grain yield in wheat. *Genes, Genomes, Genetics*, G3(4), 1231–1247. <https://doi.org/10.1101/389825>
- Lado, B., Barrios, P., Quincke, M., Silva, P., & Gutiérrez, L. (2016). Modeling genotype × environment interaction for genomic selection with unbalanced data from a wheat breeding program. *Crop Science*, 56(5), 2165–2179. <https://doi.org/10.2135/cropsci2015.04.0207>
- Luo, R., & Qi, X. (2017). Function-on-function linear regression by signal compression. *Journal of the American Statistical Association*, 112(518), 690–705. <https://doi.org/10.1080/01621459.2016.1164053>
- Luvisi, A., Ampatzidis, Y., & De Bellis, L. (2016). Plant pathology and information technology: Opportunity for management of disease outbreak and applications in regulation frameworks. *Sustainability*, 8(8), 831. <https://doi.org/10.3390/su8080831>
- Midmore, D., Cartwright, P., & Fischer, R. (1982). Wheat in tropical environments. I. Phasic development and spike size. *Field Crops Research*, 5, 185–200. [https://doi.org/10.1016/0378-4290\(82\)90022-3](https://doi.org/10.1016/0378-4290(82)90022-3)
- Montesinos-López, A., Montesinos-López, O., Cuevas, J., Mata-López, W., Burgueño, J., Mondal, S., et al. (2017a). Genomic Bayesian functional regression models with interactions for predicting wheat grain yield using hyper-spectral image data. *Plant Methods*, 13(1), 1–29. <https://doi.org/10.1186/s13007-017-0212-4>
- Montesinos-López, O., Montesinos-López, A., Crossa, J., De Los Campos, G., Alvarado, G., Suchismita, M., et al. (2017). Predicting grain yield using canopy hyperspectral reflectance in wheat breeding data. *Plant Methods*, 13(1), 1–23. <https://doi.org/10.1186/s13007-016-0154-2>
- Ortiz, R., Sayre, K., Govaerts, B., Gupta, R., Subbarao, G., Ban, T., et al. (2008). Climate change: can wheat beat the heat? *Agriculture, Ecosystems & Environment*, 126(1–2), 46–58. <https://doi.org/10.1016/j.agee.2008.01.019>
- Pantazi, X., Moshou, D., Alexandridis, T., Whetton, R., & Mouazen, A. (2016). Wheat yield prediction using machine learning and advanced sensing techniques. *Computers and Electronics in Agriculture*, 121, 57–65. <https://doi.org/10.1016/j.compag.2015.11.018>
- Pradhan, S., Babar, M., Robbins, K., Guo, J., Bai, G., Mason, R., et al. (2019). Understanding the Genetic Basis of Spike Fertility to Improve Grain Number, Harvest Index, and Grain Yield in Wheat Under High Temperature Stress Environments. *Frontiers in Plant Science*, 10, 1481. <https://doi.org/10.3389/fpls.2019.01481>
- Ranganathan, P., Pramesh, C., & Aggarwal, R. (2017). Common pitfalls in statistical analysis: logistic regression. *Perspectives in Clinical Research*, 8(3), 148. https://doi.org/10.4103/picr.PICR_87_17
- Raun, W., Solie, J., Johnson, G., Stone, M., Lukina, E., Thomason, W., et al. (2001). In-season prediction of potential grain yield in winter wheat using canopy reflectance. *Agronomy Journal*, 93(1), 131–138. <https://doi.org/10.2134/agronj2001.931131x>
- Reynolds, M., Pask, A., & Mullan, D. (2012). Physiological breeding I: interdisciplinary approaches to improve crop adaptation. (Texcoco, Mexico): *CIMMYT*. <http://hdl.handle.net/10883/1287>
- Reynolds, M., Singh, R., Ibrahim, A., Ageeb, O., Larqué-Saavedra, A., & Quick, J. (1998). Evaluating physiological traits to complement empirical selection for wheat in warm environments. *Euphytica*, 100, 84–95. https://doi.org/10.1007/978-94-011-4896-2_20

- Rutkoski, J., Poland, J., Mondal, S., Autrique, E., González Pérez, L., Crossa, J., et al. (2016). Canopy temperature and vegetation indices from high-throughput phenotyping Improve accuracy of pedigree and genomic selection for grain yield in wheat. *Genes GeneticsGenomics*, 6, 2799–2808. <https://doi.org/10.1534/g3.116.032888>.
- Ruß, G., Kruse, R., Schneider, M., & Wagner, P. (2008). Data mining with neural networks for wheat yield prediction. In P. Perner (Ed.), *Industrial Conference on Data Mining* (pp. 7–56). Berlin, Heidelberg: Springer.
- Simmons, S., Oelke, E., & Anderson, P. (1985). *Growth and development guide for spring wheat*. University of Minnesota Agricultural Extension Service: AG-FO-2547
- Talukder, S., Babar, M., Vijayalakshmi, K., Poland, J., Prasad, P., Bowden, R., et al. (2014). Mapping QTL for the traits associated with heat tolerance in wheat (*Triticum aestivum* L.). *BMC Genetics*, 15(1), 1–13. <https://doi.org/10.1186/s12863-014-0097-4>.
- Trnka, M., Rötter, R., Ruiz-Ramos, M., Kersebaum, K., Olesen, J., Žalud, Z., et al. (2014). Adverse weather conditions for European wheat production will become more frequent with climate change. *Nature Climate Change*, 4(7), 637–643. <https://doi.org/10.1038/nclimate2242>
- Wardlaw, I., Sofield, I., & Cartwright, P. (1980). Factors limiting the rate of dry matter accumulation in the grain of wheat grown at high temperature. *Functional Plant Biology*, 7(4), 387–400. <https://doi.org/10.1071/pp9800387>
- Watson, P., & Petrie, A. (2010). Method agreement analysis: a review of correct methodology. *Theriogenology*, 73(9), 1167–1179. <https://doi.org/10.1016/j.theriogenology.2010.01.003>
- Zhang, L., Zhang, Z., Luo, Y., Cao, J., & Tao, F. (2020). Combining optical, fluorescence, thermal satellite, and environmental data to predict county-level maize yield in China using machine learning approaches. *Remote Sensing*, 12(1), 21. <https://doi.org/10.3390/rs12010021>
- Zhang, Y., Qin, Q., Ren, H., Sun, Y., Li, M., Zhang, T., et al. (2018). Optimal hyperspectral characteristics determination for winter wheat yield prediction. *Remote Sensing*, 10(12), 2015.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Lucas Costa¹ · Jordan McBreen² · Yiannis Ampatzidis¹ · Jia Guo² · Mostafa Reisi Gahrooei³ · Md Ali Babar²

¹ Department of Agricultural and Biological Engineering, Southwest Florida Research and Education Center, University of Florida, IFAS, 2685 SR 29 North, Immokalee, FL 34142 Gainesville, USA

² Department of Agronomy, University of Florida, IFAS, Main Campus, 3105 McCarty Hall B, Gainesville, FL 32611, USA

³ Department of Industrial and Systems Engineering, University of Florida, Main Campus, 303 Weil Hall, Gainesville, FL 32611, USA