

ORIGINAL RESEARCH

Phenomic selection and prediction of maize grain yield from near-infrared reflectance spectroscopy of kernels

Holly M. Lane¹  | Seth C. Murray¹  | Osvaal A. Montesinos-López²  |
Abelardo Montesinos-López³ | José Crossa⁴  | David K. Rooney¹ |
Ivan D. Barrero-Farfan¹ | Gerald N. De La Fuente¹  | Cristine L. S. Morgan^{1,5} 

¹Department of Soil and Crop Sciences, Texas A&M Univ., College Station, TX 77843-2474, USA

²Facultad de Telemática, Universidad de Colima, Colima, Colima, 28040, México

³Departamento de Matemáticas, Centro Universitario de Ciencias Exactas e Ingenierías (CUCEI), Universidad de Guadalajara, Guadalajara, Jalisco 44430, México

⁴Biometrics and Statistics Unit, International Maize and Wheat Improvement Center (CIMMYT), México City, Mexico

⁵Current address: Soil Health Institute, 2803 Slater Road, Suite 115, Morrisville, NC 27560, USA

Correspondence

Seth C. Murray, Dep. of Soil and Crop Sciences, Texas A&M Univ., College Station, TX 77843-2474, USA.

Email: sethmurray@tamu.edu

Funding information

Texas Corn Producers Board; Texas A&M University, Grant/Award Numbers: College of Agriculture and Life Sciences Excellence, Eugene Butler Endowed Chair; National Institute of Food and Agriculture, Grant/Award Numbers: 2010-85117-20539, 2017-67013-26185, Hatch funds; Texas A&M University Agrilife Research

Abstract

High-throughput phenotyping technologies, which can generate large volumes of data at low costs, may be used to indirectly predict yield. We explore this concept, using high-throughput phenotype information from Fourier transformed near-infrared reflectance spectroscopy (NIRS) of harvested kernels to predict parental grain yield in maize (*Zea mays* L.), and demonstrate a proof of concept for phenomic-based models in maize breeding. A dataset of 2,563 whole-kernel samples from a diversity panel of 346 hybrid testcrosses were scanned on a plot basis using NIRS. Scans consisted of 3,076 wavenumbers (bands) in the range of 4,000–10,000 cm⁻¹. Corresponding grain yield for each sample was used to train phenomic prediction and selection models using three types of statistical learning: (a) partial least square regression (PLSR), (b) NIRS best linear unbiased predictor (NIRS BLUP), and (c) functional regression. Our results found that NIRS data were a useful tool to predict maize grain yield and showed promising results for evaluating genetically independent breeding populations. All model types were successful; functional regression followed by the PLSR model resulted in the best predictions. Pearson's correlations between predicted and observed grain yields exceeded .7 in many cases within random cross validation.

Abbreviations: AF, aflatoxin; BLUE, best linear unbiased estimator; BLUP, best linear unbiased predictor; CV, cross validation; CV0, predicting one environment using data from all other environments; CV1, 20% of the hybrids are predicted by the remaining 80% of hybrids (five-fold), within each environment; CV2, predicting across environments, where hybrids are seen in some environments but predicted in others (mimics sparse testing); G × E, genotype by environment; G-BLUP, genomic best linear unbiased predictor; GEM, germplasm enhancement of maize lines; GWAS, genome-wide association study; LM, simple linear model; NIRS, near-infrared reflectance spectroscopy; NIRS BLUP, NIRS-based best linear unbiased predictor; PLSR, partial least squares regression; RMSEP, root mean square error of prediction; SERAT, southeast regional aflatoxin trial; UAS, unoccupied aerial systems; WS, water stress, unirrigated treatment; WW, well-watered, irrigated treatment.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs 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.

© 2020 The Authors. *The Plant Phenome Journal* published by Wiley Periodicals, Inc. on behalf of American Society of Agronomy and Crop Science Society of America.

Partial least squares regression also showed promise on independent breeding trials. More research on predicting phenotypic traits from spectra will provide better understanding how NIRS and other phenomic technology can be used in predicting phenotypes of breeding programs.

1 | INTRODUCTION

1.1 | Advancement of plant breeding

Modern plant breeding is conducted much differently from the crop domestication of over 10,000 yr ago, and even from the first scientific breeding of 100 yr ago (Lee & DeVore, 1968). As new science for plant breeding has progressed, a major focus has been on genomic tools for crop improvement, with key technological innovations being the primary driver of advancements. Over time, breeding has become faster, by decreasing cycle times (i.e., growing plants in off-season nurseries to reach multiple generations within a year), and has become more targeted, through marker-assisted selection and other molecular techniques (Brummer et al., 2011). Most recently, genomic prediction and selection approaches have shown practical value and have become extensively relied on for decision making in private and public sector breeding programs (Cooper et al., 2004; Crossa et al., 2017; Zhao, Mette, & Reif, 2015). In general, technologies to better understand and collect genetic data have vastly improved over the past few decades (Levy & Myers, 2016), leading to other bottlenecks in the breeding process.

In contrast, methods employed to collect and quantify phenotypic data have not seen the same rate of technological advancement (Araus & Cairns, 2014). It has been noted that phenotyping is the limiting factor in genomic selection and other breeding methods (Rincent et al., 2012). Evaluating plants in the field is the single most costly and complex aspect of a breeding program, in terms of time, labor, and money, but is a key step in most accurately estimating a genotype's yield potential and training genomic models (Bernardo, 2008; Bernardo & Yu, 2007). In addressing phenotyping bottlenecks, efforts are expanding to better collect and analyze phenotypic data on a wide scale. Recent findings suggest that improving phenotyping methods may not just simply improve genomic selection or other genetics-based techniques, but could potentially reduce or eliminate the need for genotyping in some cases (Rincent et al., 2018).

1.2 | Near-infrared reflectance spectroscopy

Near-infrared reflectance spectroscopy (NIRS) has a long history in agricultural settings for evaluating chemical

composition since the 1960s, with some of the first mentions of the technology being used to quantify seed oil composition (Morris & Holman, 1961). Near-infrared reflectance spectroscopy is now commonly used to estimate characteristics of grain or seeds including protein, starch, and oil content (Chen, Ren, Zhang, Diao, & Shen, 2013; Ferreira, Pallone, & Poppi, 2013; Murray et al., 2008a; Silva, Milach, Silva, & Montero, 2008) and new composition traits such as grain phenolics (Meng et al., 2015). Near-infrared reflectance spectroscopy composition estimates are commonly made on plant tissue, such as forage and silage, to estimate bioenergy potential or livestock feeding value (Murray et al., 2008b; Wolfrum et al., 2013). Moreover, NIRS has a sound grounding in physics, as near-infrared light at specific wavelengths is absorbed by specific chemical bonds that make up the components of living tissues. Furthermore, the relative proportion of each of these bonds within the tissue quantitatively influence the nature of the absorbance of light at different wavenumbers (Foley et al., 1998). Typically, however, no single region of the near-infrared spectrum can be attributed to a single chemical component in plant and animal tissues, because they exist as heterogeneous mixtures of compounds, making direct inference from spectra challenging (Foley et al., 1998). Statistical analysis, traditionally partial least squares regression, is needed to extract meaningful information from NIRS on biological samples. Key among the benefits of NIRS is that there is no need to contact, damage, or destroy samples. Consequently, samples can be analyzed and still be available for planting or other analyses.

Although previous studies have shown the ability of NIRS to quantify composition, understand chemical bonds, differentiate between species, varieties, or even genetics (Bertrand, Robert, & Loisel, 1985; Espinoza, Hodge, & Dvorak, 2012; Lang, Almeida, & Costa, 2017), using NIRS information to perform selections in the context of breeding for grain yield remains relatively novel, since grain yield is not a chemical component. However, NIRS has been previously shown to correlate with yield. Ferrio, Bertran, Nachit, Català, and Araus (2004) investigated grain yield correlation with NIRS of durum wheat (*Triticum aestivum* L.) flour. Despite high r values, the slopes of their predicted and measured yield values were not in unity, causing them to conclude that the ability of NIRS to predict grain yield in wheat was not an accurate way to achieve estimations under their conditions. They partially

attributed these challenges to a lack of strong previous calibrations under the same growing conditions of the samples.

1.3 | Phenomic prediction and selection for grain yield from grain or plant spectra

The concept of treating NIRS data as markers, and using them to make inferences about relatedness, is a recent phenomenon. The theory and the term of phenomic selection was first introduced by Rincent et al. (2018), as the practice of evaluating and selecting lines via phenotypic variables. Essentially, phenotypic, or phenomic, variables (such as NIRS) replace genetic marker information in traditional selection methodology. Rincent et al. (2018) used NIRS wavelength data on wheat grain and leaf tissue as well as poplar (*Populus nigra* L.) wood to develop NIRS best linear unbiased predictions (NIRS best linear unbiased predictor [NIRS BLUP], an NIRS similarity matrix); the authors surprisingly obtained better prediction accuracy than using genomic best linear unbiased predictor (G-BLUP; genomic similarity matrix). The present study serves to independently validate and generalize some of the concepts and findings of Rincent et al. (2018) about the predictive capacity of NIRS, using phenomic prediction (plot basis) and phenomic selection (entry basis).

The main objective of this study was to apply models and methods to the NIRS data for the phenomic prediction and selection of grain yield. Specifically, our aim was to predict parental grain yield from existing grain scans of a hybrid maize (*Zea mays* L.) diversity panel grown in several environments (Barrero-Farfan et al., 2015; Flint-Garcia et al., 2005; Warburton et al., 2013) using cross validation and apply these calibrations on additional independent breeding populations to validate calibration robustness.

We sought to evaluate different statistical models: (a) partial least square regression (PLSR); (b) NIRS BLUP, employing the values of NIRS instead of markers as proposed by Rincent et al. (2018); and (c) functional regression analyses using the NIRS values as covariates in a $G \times E$, multi-environment model. In this case, PLSR was used as a *phenomic prediction*, where yield is predicted on a plot sample basis, with NIRS used for each plot. This is analogous to how composition of grain samples have traditionally been predicted. The NIRS BLUP and functional regression models were used as *phenomic selection*, where predictions were made on an entry basis, with NIRS being averaged within pedigrees. This is analogous to how a breeding program predicts breeding values using marker information.

Different cross-validations schemes were used to ensure prediction accuracy estimates would be relevant to future breeding activities. These schemes addressed (a) predicting known genotypes in unknown environments, (b) predicting unknown genotypes in known environments, and (c) so-called

Core Ideas

- Phenomic selection and prediction using near-infrared reflectance spectroscopy (NIRS) of maize kernels was successful.
- Two methods of implementing phenomic selection in maize showed promise.
- Predicting yield on an entry and plot basis have potential in maize breeding.
- Grain NIRS likely extends beyond traditional settings or uses.

sparse testing, where known genotypes are predicted in known environments in which they are not observed. If NIRS can be used to estimate the performance (grain yield) of a genotype, either on a plot or entry basis, it therefore stands to reason that the spectra also capture some of the value of that material.

2 | MATERIALS AND METHODS

2.1 | Experimental design and germplasm

A set of 346 diverse hybrid lines were grown as a structured genome-wide association study (GWAS) to assess aflatoxin resistance, drought tolerance, and other agronomic traits such as yield in 2011 and 2012 in College Station, TX (Barrero-Farfan et al., 2015). These lines originated from a subset of the USDA–Goodman maize association panel (Flint-Garcia et al., 2005), as well as the southern subtropical Williams/Warburton panel (Warburton et al., 2013). This panel was crossed to two isogenic lines of Tx714, a high-yielding, southern United States bred stiff-stalk line that is more than 95% identical to its stiff-stalk relative, B73 (Romay et al., 2013). The Tx714 isogenic hybrids differed only for which of two maize lipoxygenase genes were mutated (De La Fuente et al., 2013; Park, Kunze, Ni, Feussner, & Kolomiets, 2010). Each isogenic hybrid was grown under one or two experimental conditions (well-watered [WW] and limited irrigation, or water stress [WS]) with two replicates in a randomized complete block design, where seed was available. A full description of the hybrids and experimental design is in Barrero-Farfan et al. (2015). Since isogenic hybrids did not differ in grain yield, the data from both were combined and treated as the same hybrid. This population was used for all model training and for validation.

Other elite hybrid breeding trials in the Texas A&M breeding program were also used as independent validation sets for evaluating prediction robustness. These trials, grown in 2011, consisted of breeding relevant hybrids with no known or expected relatedness to the original material (GWAS

hybrids) presented above, most being sub-tropical derived lines crossed to U.S. commercial stiff stalk hybrids (Murray et al., 2019). In addition to grain yield, these tests were grown to assess aflatoxin (AF) resistance in the hybrids. Material from within the program represented four tests, and are referred to in this study as 1AF, 2AF, 3AF, and 4AF. Two other tests, which included breeding material from other programs, were also assessed in 2011. The Southeast regional aflatoxin trials (SERAT; Wahl et al., 2017), and the germplasm enhancement of maize (GEM) lines.

Three of these validation tests (3AF, SERAT, and GEM) were grown in College Station, TX, under similar conditions, 1AF and 2AF were grown in both Weslaco and College Station, TX, and 4AF was grown in Corpus Christi, TX. All together, these datasets combined represent 200 pedigrees across three Texas locations from 2011 (679 samples total), most of which were breeding material of commercial hybrid checks. Grain yield and NIRS were collected for each breeding test on a plot basis. This set of tests was used as a practical validation for how broadly the trained models could be used in a breeding program.

2.2 | Spectral data

Near-infrared reflectance spectroscopy was performed on harvested grain from Barrero-Farfan et al. (2015) hybrids during the completion of the initial study. In total, 2,563 whole kernel samples, each bulked from a single plot, were scanned using Fourier transformed near-infrared reflectance spectroscopy. Spectra were collected with a Thermo Anteris II Fourier transformed interferometer, where approximately 175 g of bulked grain, acclimated to a climate-controlled room, were scanned 128 times, resulting in an average (full description of methodology in Meng et al., 2015). The spectra consist of every other wavenumber (i.e., band) from 4,000 to 10,000 cm^{-1} (a range of 1,000–2,500 nm). The raw spectral data are included in Supplement 1 and 2.

2.3 | Data processing

In 2011, very low yields were observed in all experiments due to record heat and drought. The lowest yielding of the plots appeared remarkably atypical, so a threshold of at least 62 g m^{-2} (10 bu ac^{-1}) per plot was set. Applying the threshold, 47 plots were removed, leaving 2,516 samples with an average yield of 559 g m^{-2} , and a range of 63–1,462 g m^{-2} . Furthermore, as an objective method to eliminate bad NIRS scans, an algorithm was implemented in R (R Core Team, 2018) to eliminate scans that had shapes perceptively different from other scans. First, the difference of each spectra from the average spectra across each wavenumber was calculated.

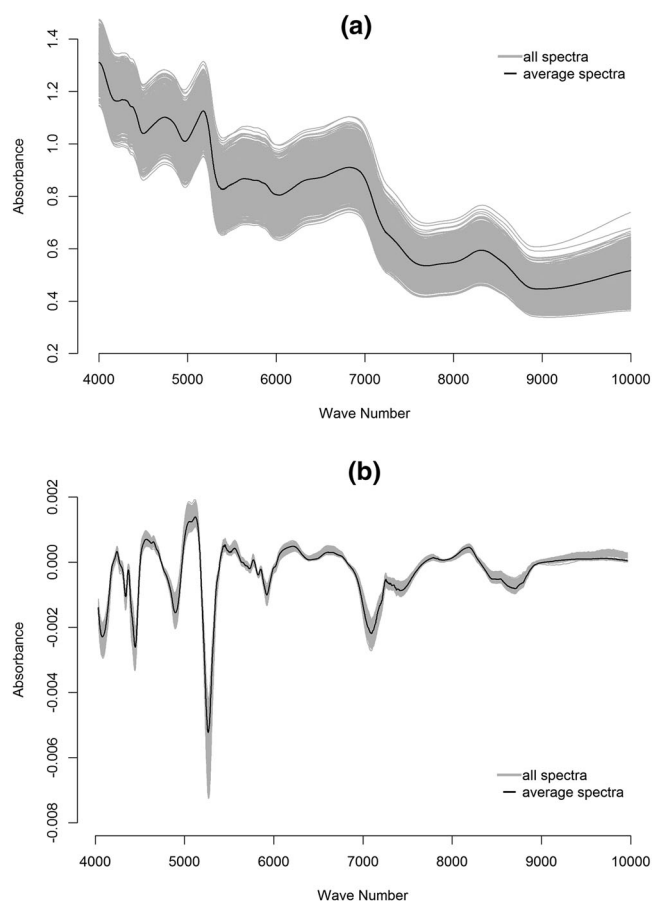


FIGURE 1 All analyzed spectra from the set of whole kernel scans are shown in grey for (a) untreated scans, or raw spectra and (b) the Savitsky-Golay first derivative of the scans. The black line in each graph is the average of all scans

Next, the sum of the absolute value of those differences were taken. Finally, scans with substantial differences from the average were removed. This method successfully and objectively eliminated spectra with abnormal shapes (“bad scans”), as opposed to abnormal predictions within the model. This, presumably, was an unbiased approach to identifying bad scans. This procedure removed 24 sample scans leaving 2,492 samples for final analyses. Supplement 3 contains this filtered, treated data.

The cleaned scans were loaded into R and processed with the *prospectr* package (Stevens & Ramirez-Lopez, 2014) to obtain the first and second derivatives, using the Savitsky-Golay method and a window size of 37 (Figure 1). This transformation served as a normalization from the raw scans, and mainly accounted for differences in albedo (overall reflectance) between samples. Interferometers, as used in this study, generally have a higher spectral resolution than other methods, often associated with a smaller signal/noise ratio. Therefore, it has been recommended to use smoothing derivation methods such as the Savitsky-Golay method as opposed to a simpler, finite-difference derivative method (Rinnan, van

den Berg, & Engelsens, 2009). Ultimately, the first derivative proved to offer the best results in cross validation, so this was used across all model types.

2.4 | Statistical models to predict grain yield using NIRS information

Multiple statistical approaches were investigated for grain yield prediction accuracy. Both phenomic prediction models, where data were assessed on a plot basis, and phenomic selection models, where data was assessed on an entry basis, are described. The models using the NIRS and phenotypic information are described below.

2.4.1 | Partial least squares regression

Partial least squares regression was used to predict yield as implemented in the `pls` package in R (Mevik & Wehrens, 2007). Models were cross-validated using the default of 10 segments, and then further validated on a held-out test set. Held-out test sets were selected to ensure that entire instances of a hybrid were held out from all environments, in contrast to `pls` defaults for random cross validation. Held-out test sets were selected using the `DUPLEX` method available in the `prospectr` package (Stevens & Ramirez-Lopez, 2014), which functions by selecting the two most distant points (based on principal components) iteratively for the sample size desired. This served to select a spectrally representative set of samples to use for validation or calibration while holding out all replicates of that hybrid for training the model; this is similar to predicting untested genotypes in known environments (Saint Pierre et al., 2016).

For example, for the model trained on both years (PLSR 1, Table 1 in Results), the training set consisted of 1,573 samples, representing 253 pedigrees. The test set was composed of 848 samples, representing 91 pedigrees. All PLSR models were validated in this way, though the number of samples used to train and test the models varied.

Additionally, PLSR models were built to investigate the ability for the model to be trained on one year to predict the same year, or to predict the other year. Models were evaluated either for their ability to predict known or unknown hybrids in an unknown year. Evaluating the ability of the model to predict yield for a separate experiment from another year, consisting of different hybrids, or predicting unknown genotypes in unknown environments (Saint Pierre et al., 2016), is also of great practical interest in breeding programs.

The number of PLSR components used in each model were selected based on the root mean squared error prediction (RMSEP) curve using the one-sigma heuristic and per-

mutation approaches as implemented in the `pls` package in R (Mevik & Wehrens, 2007); R code used for PLSR models can be found in Supplement 4.

2.4.2 | Baseline single environment and across environment models for implementing NIRS BLUP and functional regression

In this subsection, we explain the six phenomic selection models implemented using the phenotypic and NIRS information under NIRS BLUP and functional regression approaches. Before describing the six models used under BLUP and functional regression, we briefly describe the model used for obtaining the BLUEs (best linear unbiased estimators) for grain yield and for NIRS over each of the wavenumbers with the intention of removing the experimental design. This analysis was the first stage analyses that produced the BLUEs used in the second stage analysis where Models M1 to M6 are implemented. The BLUEs of each hybrid and NIRS in each environment were obtained using the following model:

$$y_{jk} = \mu_j + b_k + H_j + \varepsilon_{jk} \quad (1)$$

where y_{jk} is the response variable (grain yield, or wavenumber measurements) of the j th hybrid in the k th complete block; μ is the overall mean; b_k is the random effect of the k th complete block assuming independent and identically distributed (iid) $N(0, \sigma_b^2)$; H_j is the fixed effect of the j th hybrid, and ε_{jk} , assuming iid $N(0, \sigma_\varepsilon^2)$, represents the random residual plot error associated with the observation y_{jk} .

Then in the second stage analysis these BLUEs of grain yield were used as response variable and the BLUEs of each NIRS were incorporated as covariates to implement an improved version of the following model that takes into account the hybrid \times environment interaction

$$y_{ij} = \mu + E_i + H_j + HE_{ij} + \varepsilon_{ij} \quad (2)$$

where E_i is the fixed main effect of the i th environment; H_j is main random effect of the j th hybrid identical and independent normal distributed (iid) $HE_{ij} \sim N(0, \sigma_{HE}^2)$; HE_{ij} is the random interaction effect between the i th environment and the j th hybrid; it is assumed iid $N(0, \sigma_{HE}^2)$; and ε_{ij} is the residual error term assuming iid $N(0, \sigma^2)$. For the sake of clarity, later we named the model including hybrid \times environment interactions as Model M5.

2.4.3 | The NIRS BLUP model with interaction: M1

Wavenumber can be introduced in the previous baseline model such that the effect of hybrid (H_j) can be replaced by w_j ,

TABLE 1 Description of factors used in training partial least squares regression (PLSR) and simple linear model (LM) models for grain yield on all genome-wide association study (GWAS) data and their results. The number of samples used in the training and testing sets as well as the number of genotypes (pedigrees) represented by those samples are reported. The number of partial least squares components used for each model are listed (PLS) when relevant. Models were built with 10-fold cross validation and then applied to the testing set. Pearson's correlation (r) and root mean square error of prediction (RMSEP) in g m^{-2} were calculated between observed and predicted values of the test set. Partial least squares regression 1 and 2 are built using information from both 2011 and 2012

Scenario	Training data		Testing data		Predictor	PLS	r	RMSEP
	Samples	Pedigrees	Samples	Pedigrees				
PLSR 1	1,573	253	848	91	NIRS	15	.84	163.67
PLSR 2 ^a	1,389	254	734	90	NIRS	11	.82	174.52
LM 1	1,389	254	734	90	Protein ^b	–	.59	247.25
LM 2	1,389	254	734	90	Starch ^b	–	.21	296.38
LM 3	1,389	254	734	90	Oil ^b	–	.12	301.86
LM 4	1,389	254	734	90	Protein/starch/oil ^b	–	.63	238.52

^aThe full data set did not have corresponding composition estimations for each sample, so a baseline PLSR was run with the subset of the data for which composition was available.

^bProtein, starch, and oil values were estimations based on previously trained NIRS models calibrated with wet chemistry.

which is expressed as a linear regression of phenotypic data on the wavenumber covariates that approximates the value of the j th hybrid such that the vector of the NIRS random effect is assumed $\mathbf{w} \sim N(0, \mathbf{W}\sigma_w^2)$, where σ_w^2 is the spectral variance component and \mathbf{W} is a scaled wavenumber relationship matrix that was computed using the matrix of wavenumber information (\mathbf{N}) from data of order $J \times w$ (for a total of J lines and w wavenumbers) as $\mathbf{W} = \mathbf{N}\mathbf{N}^T/w$.

It is important to point out that the final \mathbf{W} reported for lines was calculated as the average of the wavenumber relationship matrices calculated in each environment since the wavenumber reflectance values were obtained for each hybrid in each environment. Similarly, for the interaction terms, when wavenumber information was used, the hybrid \times environment interaction HE_{ij} of the lines across environments baseline model was replaced by wE_{ij} the random effect of the interaction term of the i th environment and the j th wavenumber and $\mathbf{wE} = [wE_{11}, \dots, wE_{IJ}]^T \sim N(0, \mathbf{W}_A\sigma_{wE}^2)$, where σ_{wE}^2 is the variance component associated with the wavenumber \times environment interaction, but \mathbf{W}_A was calculated as

$$\mathbf{W}_A = \frac{\mathbf{N}_A\mathbf{N}_A^T}{w}$$

where \mathbf{N}_A is the matrix of wavenumber information of order $JI \times w$ where I denotes the number of environments assuming that the same number of hybrids, J , were evaluated in each environment.

Therefore, using basic wavenumber information, the NIRS BLUP model including environments, wavenumber, and their interactions becomes Model M1, which is

$$y_{ij} = E_i + w_j + wE_{ij} + \varepsilon_{ij} \quad (3)$$

M1 establishes relationships between hybrids based on wavenumber information, which is similar to the conventional

G-BLUP. In a traditional G-BLUP approach, dense molecular markers are used (instead of wavenumbers), and covariates are incorporated to develop the genomic relationship matrix (\mathbf{G}) establishing the similarities between hybrids based on genomic markers. The main difference in this scenario is that that wavenumbers are measured for all hybrids in each environment, as opposed to DNA markers, which are measured once for each hybrid. The R code used for NIRS BLUP (M1) can be found in Supplement 4.

2.4.4 | Functional regression model with interaction: M2

High throughput phenotyping technologies, in addition to NIRS, are increasingly available in agriculture and generate large volumes of data. Many such phenotypes are multi-dimensional, and these data can be represented as functions. Functional data analysis is a field of study that deals with the analysis and theory of data whose units of observation are functions (curves) defined as continuous domain (Morris, 2015). Recent studies by Montesinos-López et al. (2017a, 2017b) present functional regression models to develop prediction equations for wheat grain yield and other traits using hyperspectral crop image data from the field. This work showed that functional regression models could provide yield predictions with similar and, in some cases, higher predictive power than that of conventional regression techniques. Montesinos-Lopez et al. (2018) described details for implementing Bayesian functional regression using the developed genomic functional regression package. The previously defined baseline model, $y_{ij} = E_i + H_j + HE_{ij} + \varepsilon_{ij}$, can be combined with the information from the wavenumber, but not as a wavenumber matrix as introduced in Model M1. Instead, wavenumber values are combined using models

displaying wavenumbers as functional covariates. These models are called functional regression models, where the response variable is a scalar and some of the covariates are functions. Thus, this combination will be named Model M2 and is represented as

$$y_{ij} = E_i + H_j + HE_{ij} + \int_{4,000 \text{ cm}^{-1}}^{10,000 \text{ cm}^{-1}} w_{ij}(k) \beta_1(k) dk + \varepsilon_{ij} \quad (4)$$

Here, $w_{ij}(k)$ is the functional predictor and represents the value of a continuous underlying process evaluated at the k th wavenumber; $\beta_1(k)$ is the functional regression beta coefficient for the functional part of the model; and ε_{ij} is defined as in Equation 2 (Montesinos-López et al., 2017b).

The functional regression models add the 3,076 wavenumbers to M2 as functional covariates constructed over the interval between 4,000 and 10,000 cm^{-1} . Model M2 was implemented using a total number of $L = 101$, Fourier basis, to approximate the functions, $w_{ij}(k)$ and $\beta_1(k)$ in Model M2. The Fourier basis (not to be confused with Fourier transformed NIRS data collection) is a set of predefined functions in terms of sine and cosine. If in Model M2, we replace the integral of the product by the sum of products ($\sum_{k=1}^p w_{ijk} \beta_k$), then Model M2 is reduced to a conventional linear regression model. In this model, w_{ijk} represents the k th wavenumber data measured on the j th hybrid in the i th environment with $k = 1, 2, \dots, 3,076$; w_{ijk} are BLUEs obtained in the first stage analysis; and β_k is the beta regression coefficient for the k th wavenumber. Note that we implemented Model M2 to avoid implementing a model with 3,076 beta coefficients corresponding to each of the wavenumbers. This model was implemented using a Bayesian approach. The R code used for Model M2 can be found in Supplement 4.

2.4.5 | The NIRS BLUP model with only main effects of hybrids: M3

Model M3 is similar to M1 but only with main effects of hybrids (with wavenumbers incorporated through a covariance matrix) applied for each environment. This model is equal to

$$y_j = w_j + \varepsilon_j \quad (5)$$

In M2 Equation 4, $w \sim N(0, \mathbf{W}\sigma_w^2)$, where \mathbf{W} was calculated with the wavenumber information of each environment. Note that Model M3 is obtained from Model M1, ignoring the main effect of environment, E_i , and the interaction term, wE_{ij} . It is reduced to $y_j = w_j + \varepsilon_j$, with $w = [w_1, \dots, w_j]^T$ and $w \sim N(0, \mathbf{W}\sigma_w^2)$.

2.4.6 | Functional regression model with only main effects of hybrids: M4

Model M4 is similar to M2, but with only the main effects of hybrids, $y_j = H_j + \varepsilon_j$, because it was applied for each environment one at a time. When using the wavenumbers as covariates, Model M4 is described as

$$y_j = H_j + \int_{4,000 \text{ cm}^{-1}}^{10,000 \text{ cm}^{-1}} w_j(k) \beta_1(k) dk + \varepsilon_{ij} \quad (6)$$

2.4.7 | Models without NIRS: M5 and M6

As previously mentioned, the model of the hybrid \times environment interactions is named Model M5,

$$y_{ij} = \mu + E_i + H_j + HE_{ij} + \varepsilon_{ij} \quad (7)$$

Furthermore, when Model M5 includes only the random effect of hybrids, this yields Model M6.

$$y_j = H_j + \varepsilon_{ij} \quad (8)$$

2.5 | Assessing the models' prediction accuracy

For all validation schemes, we report the Pearson's correlation (r) and RMSEP between the observed and predicted values of grain yield for the testing set. For PLSR models, these results are reported for the held-out validation test sets as described above. For functional regression and NIRS BLUP models, we report the average of the five-folds resulting from the random partitions implemented.

In order to measure the accuracy of the functional regression and NIRS BLUP phenomic selection models, we used three cross-validation strategies on an entry basis, the first one of which was also used for PLSR phenomic prediction on a plot basis. These schemes were: (a) cross-validation predicting one environment using the other environments as the training set; (b) random cross-validation within each environment where 20% of the maize hybrids are predicted using 80% of the other hybrids as the training population; and (c) random cross-validation across environments where 20% of the lines were observed in some environments and predicted in other environments using the other 80% of lines.

The first scheme consists of the prediction of all hybrids in one environment (in turn, WW_2011, WW_2012, WS_2011, and WS_2012) using all the hybrids combined from the other three environments as the training set. This prediction is named cross-validation 0 (CV0) (predicting known genotypes in pseudo-unknown environments). In this experiment, when

the model was tested but not trained on an environment that data had been collected, the environment is considered to be pseudo-unknown. The PLSR, functional regression, and NIRS BLUP models were all assessed with the CV0 scheme, to test the models' abilities to perform in unknown environments.

The second cross-validation scheme was a random cross-validation within each environment, where a five-fold random partitioning was implemented. In each partition, 20% of the hybrids are predicted by the remaining 80% of hybrids; this is named (CV1) (unknown genotypes in known environments).

Finally, the last cross-validation strategy was a five-fold random cross-validation across all environments (CV2) where 20% of the hybrids were predicted by 80% of the other hybrids, similar to CV1. However, in this case, 20% of hybrids are not observed in some environments, and thus predicted in those environments, but are observed in other environments (known genotypes in known environments). This CV2 mimics a prediction problem faced by breeders in incomplete field trials where hybrids are evaluated in some, but not all, target environments (usually called *sparse testing*). Note that in CV2, all environments were used in training the model. In this cross validation, some hybrids can never be part of the training set. This is because some hybrids were not observed in all environments. This CV2 strategy used sampling with replacement, which means that one observation can appear in more than one partition. Training and testing partitions were obtained as follows: since the total number of records per trait available for the data set with multi-environments is $N = J \times I$ observations comprising of J hybrids and I environments, to select lines in the test data set, we fixed the percentage of data to be used for test (Testing = 20%). Then we chose $0.20 \times N$ (hybrids) at random, and subsequently, one environment per hybrid was randomly picked from I environments. The resulting cells (ij) were assigned to the test data set, while cells not selected through this algorithm were allocated to the training data set. Hybrids were sampled without replacement if $J \geq 0.20 \times N$, and with replacement otherwise (López-Cruz et al., 2015).

2.6 | Elite hybrid trials as validation

The breeding trials described above (1AF, 2AF, 3AF, 4AF, GEM, and SERAT) were used as practical validation and served as a secondary held out test set. In this scenario, the phenomic prediction PLSR model was trained on the original diversity panel data and then used to predict these breeding trials, as a held-out validation.

For phenomic selection Models M1–M6, all of the diversity panel data, along with 10% of the material from each test, were included in the training set. As these models function on an

entry basis, including some of the breeding trial observations was necessary in referencing the new site. Means and standard deviations across 10 repetitions of a five-fold CV are reported (see Supplement 5).

2.7 | Other analysis

To investigate if yield was simply correlating with some compositional aspect of the kernels, predictions for crude protein, starch, and oil content were obtained using existing calibrations from these components built within the research program (Christman, 2017; Meng et al., 2015). To assess the correlation of composition to yield, simple linear models (LM) were built using the LM function in R (R Core Team, 2018). These LM were compared with the results of the PLSR model trained and tested on the same samples sets, as both predict on a plot basis. Repeatability for grain yield on an entry basis across all environments were estimated as explained by Anderson, Mahan, Murray, and Klein (2018), where genetic variance was divided by error variance over the number of reps by treatments, $G \times E$ variance over treatments, and genetic variance.

3 | RESULTS

The results are presented in two main sections. The first section presents the predictions under the PLSR and LM methods. The second section describes the predictions based on the phenomic selection models (NIRS BLUP and functional regression), for single environment and $G \times E$ multi-environment approaches.

3.1 | Prediction accuracy of PLSR and LM

Cross validated yield predictions from models using PLSR within the same years demonstrated a high Pearson's correlation to maize grain yield on a plot basis in the testing data ($r = .84$; Table 1, PLSR 1). The model was successful in using spectra alone to predict the yield of a sample on a plot basis, evaluated by error in prediction, RMSEP. With only spectral data, the PLSR model using 15 PLS components predicted yield with a RMSEP of 163.67 g m^{-2} .

To investigate if kernel compositional traits were correlated with yield, NIRS composition predictions were used to predict yield using simple LM. A PLSR model as above was run but using only the 2,155 samples with composition predictions to build the training and testing sets (Table 1; PLSR 2); this had an r of .82. Looking at individual components, crude protein was the only measured component with a strong correlation with yield ($r = .58$) (Table 1; LM 1). Both starch

and fat had very low correlations ($r \leq .17$, Table 1; LM 2–3). Combining all compositional predictions into a model to predict yield improved predictions over protein alone, but only slightly, yielding an r of .64 (Table 1; LM 4).

Next, PLSR scenarios were developed to predict yield on samples from a year unknown to the model, on known or unknown hybrids (see Table 2). Results showed that a model built on 2011 predicted unknown hybrids in 2011 better than a 2012 model predicting unknown hybrids in 2012 (Table 2; PLSR 3–4). Results also showed that a model trained on 2011 predicted 2012 well, and better than the reverse (PLSR 5 and 8); even though the predicted hybrids were known to the model in both cases. In predictions between years, the best performing model was trained on all available samples from 2011 to predict 2012 (PLSR 6). The PLSR 6 had more training data than PLSR 5, as this training data represented all available 2011 data (including the USDA and other breeding populations from 2011). We also investigated the ability of a model to predict only unknown hybrids from a new year (PLSR 7), which was comparable to when most of the hybrids were already known from a previous year (PLSR 5) cross-validation under CV0, CV1, and CV2 schemes.

3.1.1 | Prediction of observed hybrids in new environments (CV0)

Results of cross-validation CV0, where one environment is predicted using the other three environments (year \times irrigation) as the training data, for PLSR, NIRS BLUP (M1), and functional regression (M2) models are shown in Tables 3 and 4. The PLSR had slightly higher prediction accuracies across all scenarios than both M1 and M2 ($r = .55$; Table 3). Model M2, using functional regression with wavelength as covariates, gave higher prediction on average when compared with Model M1, using NIRS BLUP ($r = .53$ vs. $.40$; Table 4). Error, as presented via the average RMSEP were nearly identical between functional regression and PLSR models (272.12 and 272.70 g m⁻², respectively). In general, all hybrids predicted in individual environments had higher prediction accuracies for Model M2 than for Model M1 (e.g., WS_2012 = 0.68 for M1 vs. WS_2012 = 0.74 for M2). Results showed that the prediction of hybrids unobserved in WW using observed hybrids in WS is better than the reverse, and this was much higher for M2 than for M1 (WS = 0.42 and WW = 0.75 for M2 vs. WS = 0.20 and WW = 0.28 for M1; Table 4). The PLSR performed similarly to M2 overall (Table 3). Similarly, hybrids grown in 2012 were more predictable than hybrids in 2011, which makes sense given 2011 was a drought year, presenting more overall stress to plants. Based on the r values between samples observed and predicted values, and the RMSEP results, functional regression Model M2 and PLSR models had substantially higher prediction accu-

racies of unobserved hybrid performance than NIRS BLUPs Model M1.

3.1.2 | Prediction of unobserved hybrids within each environment (CV1)

Results of Models M3 and M4 under CV1 for the predictions of hybrids within each environment are shown in Table 5. Models M3 and M4 are the counterpart of Models M1 and M2, respectively, but include only main effects of hybrids. In general, results showed that NIRS BLUP Model M3 with only main effects predicts the unobserved hybrids better (0.57) than the main effect functional regression Model M4 (0.46). The prediction accuracy of unobserved hybrids in WS_2012 for M3 (0.54) was lower than that achieved in WS_2012 for M4 (0.64).

Within an irrigation treatment, prediction of unobserved hybrids tested under WS using observed hybrids evaluated under WW was higher than the prediction of hybrids under WW using hybrids in WS by means of Models M3 and M4, if both years were included (Table 5). The prediction of hybrids in WS environment was of 0.85 with Model M3 and of only 0.47 for Model M4. In this case, unobserved hybrids in 2011 were more predictable than unobserved hybrids in 2012. Based on r between observed and predicted values, results indicated that functional regression Model M4 had lower prediction accuracy than NIRS BLUP Model M3. However, based on RMSEP values, results indicated that functional regression Model M4 had higher prediction accuracy than NIRS BLUP Model M3.

3.1.3 | Prediction of unobserved hybrids in some environments (sparse testing) (CV2)

Table 6 shows the random cross-validation CV2 fitting Models M1 and M2. When including the environment \times hybrid interaction the functional regression using the NIRS as covariates (M4) gave much higher prediction accuracy of the unobserved hybrids ($r = .73$; Table 6) than the NIRS BLUP ($r = .38$). Model M2 with functional regression covariates in the G \times E multi-environment modes gave very good predictions of unobserved hybrids in WS ($r = .91$; Table 6) as well as in WW ($r = .87$); on the contrary, NIRS BLUP with G \times E multi-environment (M1) gave very poor prediction of unobserved hybrids in WS ($r = .07$) and in WW ($r = .14$) (Table 6).

For Model M1, training on WW hybrids to predict unobserved hybrids in WS conditions showed very poor performance, and the reverse was only slightly better (Table 6). It was better to train on 2011 hybrids for Model M1, although the opposite was true for Model M2. Based on r values and RMSEP, results indicated that functional regression

TABLE 2 Description of factors used in training partial least squares regression (PLSR) scenarios to carry out year-to-year predictions on grain yield and their results. The number of samples used in the training and testing sets as well as the number of genotypes (pedigrees) represented by those samples are reported. The number of partial least squares components used for each model are listed (PLS). Models were built with 10-fold cross validation and then applied to the testing set. Pearson's correlation (r) and root mean square error of prediction (RMSEP) in g m^{-2} were calculated between observed and predicted values of the testing sets. Partial least squares regression Models 3–8 were built using information from within years, as indicated in the training and testing data information in the table

Scenario	Training data			Testing data			PLS	r	RMSEP
	Year	Samples	Pedigrees	Year	Samples	Pedigrees			
PLSR 3	2011	913	213	2011	363	61	12	.81	121.03
PLSR 4	2012	768	239	2012	448	85	15	.72	195.03
PLSR 5	2011	1276	275 ^a	2012	1216	324 ^a	9	.62	247.58
PLSR 6	2011 ^b	1962	475 ^a	2012	1216	324 ^a	15	.69	209.64
PLSR 7	2011	1276	275	2012	201	69	10	.60	244.38
PLSR 8	2012	1216	324 ^a	2011	1276	275 ^a	11	.19	329.79

^aIndicates the pedigrees were not completely unique between years.

^bIndicates data from separate 2011 experiments were included.

TABLE 3 Pearson's correlation (r) between predicted and observed values for environments water stress (WS) and well-watered (WW) during years 2011 and 2012 for partial least squares regression (PLSR) phenomic prediction models. The number of partial least squares components are also reported (PLS). Root mean squared error prediction (RMSEP) is reported in g m^{-2} . Cross-validation CV0 predicted one environment using the other three environments, one irrigation regime (WW, WS) using the other irrigation regime, or 1 yr (2011, 2012) using the other year

Model	Predicted environment	PLS	r	RMSEP	Validation
PLSR	WS_2011	12	.43	402.97	CV0
PLSR	WS_2012	12	.64	206.12	CV0
PLSR	WW_2011	13	.72	232.09	CV0
PLSR	WW_2012	13	.47	235.22	CV0
PLSR	WS	12	.48	343.08	CV0
PLSR	WW	10	.81 ^a	183.49 ^b	CV0
PLSR	2011	11	.19	329.79	CV0
PLSR	2012	12	.66	248.89	CV0
	Mean	–	.55	272.70	

^aModel with highest correlation (r).

^bModel with lowest RMSEP.

Model M2 had substantially higher prediction accuracy than NIRS BLUP Model M1. Models M5 and M6, which do not include the NIRS data (Table 7), gave a much lower prediction accuracy than Models M1, M2, M3, and M4 that include NIRS, for all three cross-validations schemes (CV0, CV1, and CV2).

3.2 | Predicting on elite breeding material

The PLSR 1 model (Table 1 describes training set) was applied to other breeding tests from 2011 to investigate the ability for this calibration to predict other genetic material, relevant for breeding. This model had an ability to predict not only unknown pedigrees from within the experiment (a genetics study where all hybrids share the same tester line T×714), but also showed ability in predicting genetically distinct material from different experiments grown in 2011 on a plot basis

(Figure 2). These breeding experiments were grown across three locations in the state of Texas. Among the best predicted were 4AF ($r = .68$), a low yielding, late planted diverse test grown in Corpus Christi, SERAT ($r = .53$, Figure 2b) a moderate yielding late planted test in College Station, whereas the other breeding tests from Weslaco (1AF and 2AF) and College Station (3AF) were not particularly well correlated ($r = .23$, $.25$, $.21$, respectively). In the GEM predictions (Figure 2b; $r = .33$), two plots were predicted erroneously, having a negative yield prediction. These plots were removed from plotting and analysis. It is important to note that although these GEM plots were not accurately predicted, they were also some of the lowest yielding plots from the test.

Various phenomic selection models were similarly investigated for their ability to perform on the same unrelated material, although results were not as impressive, given that 10% of the new material was included in the training set. These results, which importantly differ in that they were conducted

TABLE 4 Pearson's correlation (r) between predictive and observed value for environments water stress (WS) and well-watered (WW) during years 2011 and 2012 for phenomic selection Models M1 and M2 including interactions between wavenumber \times environment. Root-mean squared error prediction (RMSEP) is reported in g m^{-2} . Cross-validation CV0, predicted one environment using the other three environments, one irrigation regime (WW, WS) using the other irrigation regime, or 1 yr (2011, 2012) using the other year

Model	Predicted environment	r	RMSEP	Validation
M1	WS_2011	.41	1,905.22	CV0
M1	WS_2012	.68 ^a	1,533.79	CV0
M1	WW_2011	.62	101.55 ^b	CV0
M1	WW_2012	.45	1,581.57	CV0
M1	WS	.19	762.47	CV0
M1	WW	.28	3,969.53	CV0
M1	2011	.23	638.25	CV0
M1	2012	.31	340.07	CV0
	Mean	.40	1,354.06	
M2	WS_2011	.45	322.53	CV0
M2	WS_2012	.74	148.28 ^b	CV0
M2	WW_2011	.67	215.82	CV0
M2	WW_2012	.44	220.89	CV0
M2	WS	.42	395.73	CV0
M2	WW	.75 ^a	197.21	CV0
M2	2011	.13	354.82	CV0
M2	2012	.64	321.69	CV0
	Mean	.53	272.12	

^aModel with highest correlation (r).

^bModel with lowest RMSEP.

TABLE 5 Pearson's correlation (r) between predictive and observed values for environments water stress (WS) and well-watered (WW) during years 2011 and 2012 for Models M3 and M4 within only a single environment. Root-mean squared error prediction (RMSEP) is reported in g m^{-2} . CV1 predicted 20% of unobserved lines in that single environment using the remaining 80% of lines. SE $_r$ denotes standard error under r and SE $_{\text{RMSEP}}$ denotes standard error under RMSEP

Model	Predicted environment	r	SE $_r$	RMSEP	SE $_{\text{RMSEP}}$	Validation
M3	WS_2011	.28	0.07	75.97 ^b	18.6	CV1
M3	WS_2012	.54	0.05	148.45	51.12	CV1
M3	WW_2011	.63	0.02	101.04	33.14	CV1
M3	WW_2012	.30	0.06	184.56	58.31	CV1
M3	WS	.85 ^a	0.02	132.64	40.1	CV1
M3	WW	.77	0.03	170.97	57.57	CV1
M3	2011	.75	0.02	108.36	30.03	CV1
M3	2012	.37	0.02	203.56	40.28	CV1
	Mean	.57	–	140.69	–	
M4	WS_2011	.30	0.11	74.93 ^b	1.73	CV1
M4	WS_2012	.64 ^a	0.15	135.06	2.44	CV1
M4	WW_2011	.63	0.15	103.72	2.23	CV1
M4	WW_2012	.25	0.20	189.47	1.57	CV1
M4	WS	.47	0.21	109.22	3.1	CV1
M4	WW	.44	0.26	152.74	3.04	CV1
M4	2011	.47	0.21	90.47	2.99	CV1
M4	2012	.45	0.26	164.53	3.14	CV1
	Mean	.46	–	132.77	–	

^aModel with highest correlation (r).

^bModel with lowest RMSEP.

TABLE 6 Pearson's correlation (r) between predictive and observed values for environments water stress (WS) and well-watered (WW) during years 2011 and 2012 for Models M1 and M2 including interaction wavenumber \times environment. Root-mean squared error prediction (RMSEP) is reported in g m^{-2} . Random cross-validation CV2 for predicting 20% of the lines not observed in listed environments but observed in others. SE_r denotes standard error under r and SE_{RMSEP} denotes standard error under RMSEP

Model	Predicted environment	r	SE_r	RMSEP	SE_{RMSEP}	Validation
M1	WS_2011	.50	0.05	78.06 ^b	27.81	CV2
M1	WS_2012	.65	0.05	130.58	31.55	CV2
M1	WW_2011	.70 ^a	0.02	92.40	29.66	CV2
M1	WW_2012	.47	0.13	170.61	58.07	CV2
M1	WS	.07	0.09	254.60	45.87	CV2
M1	WW	.14	0.03	266.92	56.57	CV2
M1	2011	.12	0.05	166.05	48.31	CV2
M1	2012	.37	0.07	206.34	49.67	CV2
	Mean	.38	–	170.70	–	
M2	WS_2011	.48	0.15	80.1 ^b	2.51	CV2
M2	WS_2012	.74	0.10	120.52	2.08	CV2
M2	WW_2011	.71	0.10	88.08	2.01	CV2
M2	WW_2012	.52	0.10	167.28	103.16	CV2
M2	WS	.91 ^a	0.01	103.11	41.61	CV2
M2	WW	.87	0.03	134.85	69.51	CV2
M2	2011	.86	0.02	84.65	26.16	CV2
M2	2012	.75	0.07	147.87	73.97	CV2
	Mean	.73	–	115.81	–	

^aModel with highest correlation (r).

^bModel with lowest RMSEP.

on an entry basis, for phenomic selection Models M1–M6 are included in Supplement 5.

3.3 | Repeatability for yield

The repeatability of yield across and within environments are presented in Table 8. The lowest repeatability was within WS_2011 (0.55), and overall repeatability was relatively high (0.77), consistent with the heritabilities calculated in the original study (Barrero-Farfan et al., 2015).

4 | DISCUSSION

Our aims were to determine if information from NIRS of maize grain could be used to estimate grain yield in both phenomic prediction and selection scenarios. If the spectra capture the value of genetic material, there would be potential for these methods to be useful within the context of breeding programs. The models used in this experiment represented two distinct approaches to mining NIRS spectral data, phenomic prediction on a plot basis, and phenomic selection on an entry basis. Phenomic prediction in this case is similar to the traditional use of NIRS, with a biological composition basis likely underlying a correlated proxy yield estimate. Phenomic

selection is different in that it uses NIRS spectra to estimate relationships between genotypes and these relationships are used to estimate yield. The relative success of both approaches offers promise for future techniques using NIRS beyond traditional composition analysis.

The NIRS information clearly formed the basis for most of the prediction capability for each of the model types used in this study. Models not including NIRS information (Table 7) had very little correlation between predicted and observed values. Yield correlations (r values) between the BLUPs of each environment ranged from .32 to .67, suggesting some predictions were simply due to lines performing similarly across environments; this was often still substantially less than the NIRS models, however. The mode of action for the strong NIRS predictive relationships across genotypes and years is not yet well understood.

One hypothesis, most relevant to the plot-based phenomic prediction approach, is that spectra could be indicative of compositional aspects of the kernels that are correlated with yield. This was investigated by comparing composition prediction model results, and only crude protein appeared to have any significant inverse correlation with yield on its own (Table 1; LM 1). Combining all measured compositional predictions (starch, oil, protein; LM 4) into a model to predict yield improved predictions over a single compositional component alone; demonstrating that composition is relevant but

TABLE 7 Pearson's correlation (r) between predicted and observed values for water stress (WS) and well-watered (WW) environments during years 2011 and 2012 for Model M4, and for Models M5 and M6, which do not use NIRS in their predictions. Root-mean squared error prediction (RMSEP) is reported in g m^{-2} . Cross-validation, CV0, five-fold random cross-validation within one environments CV1 and five-fold random cross validation CV2 for predicting 20% of the lines not observed in some environments but observed in others. SE _{r} denotes standard error under r and SE_{RMSEP} denotes standard error under RMSEP. Models M4 and M6 consider the main effects of environments

Model	Environment	r	SE _{r}	RMSEP	SE _{RMSEP}	Validation
M4	WS_2011	.32	–	492.54	–	CV0
M4	WS_2012	.70	–	427.09	–	CV0
M4	WW_2011	.48	–	200.7 ^b	–	CV0
M4	WW_2012	.59 ^a	–	3,135.51	–	CV0
M4	WS	.28	–	4,095.16	–	CV0
M4	WW	.33	–	482.74	–	CV0
M4	2011	.26	–	1,182.65	–	CV0
M4	2012	.39	–	941.75	–	CV0
	Mean	.42	–	1,369.77	–	
M5	WS_2011	.01	0.06	78.45 ^b	17.12	CV1
M5	WS_2012	.04	0.09	175.46	62.18	CV1
M5	WW_2011	.04	0.05	128.83	45.48	CV1
M5	WW_2012	.06 ^a	0.03	194.6	60.29	CV1
M5	WS	–.10	0.04	251.91	54.39	CV1
M5	WW	–.09	0.05	268.33	49.75	CV1
M5	2011	–.09	0.05	161.69	49.89	CV1
M5	2012	–.10	0.03	220.83	37.88	CV1
	Mean	–.03	–	185.01	–	
M6	WS_2011	.34	0.05	92.28 ^b	33.68	CV2
M6	WS_2012	.66 ^a	0.03	142.67	58.21	CV2
M6	WW_2011	.46	0.08	114.52	39.88	CV2
M6	WW_2012	.52	0.07	169.42	58.32	CV2
M6	WS	–.30	0.02	263.42	48.64	CV2
M6	WW	–.27	0.03	281.41	55.38	CV2
M6	2011	–.15	0.03	176.22	37.02	CV2
M6	2012	.20	0.02	216.95	61.52	CV2
	Mean	.18	–	182.11	–	

^aModel with highest correlation (r).

^bModel with lowest RMSEP.

likely not sufficient in predicting yield. Using composition in any combination was never as accurate as using NIRS data. However, adding more (as yet unknown) compositional components could further improve these correlations. The NIRS-based yield predictions were surprisingly better than what is usually observed or expected in predictions of compositional traits (Christman, 2017), but this was a larger dataset.

Alternatively, more relevant to the genotypic-based phenomic selection approach, the spectra could be explaining genetic relatedness between samples. If the NIRS was capturing genetic relatedness alone, the phenomic selection models would have been expected to significantly outperform the PLSR phenomic prediction models, given that they were functioning on an entry basis, and this was not the case (see Table 3 vs. Table 4). Furthermore, the phenomic

selection models were more limited in predicting unrelated material than phenomic prediction models (see Figure 2 vs. Supplement 5). At best, phenomic selection models were comparable to phenomic prediction, never significantly better and often much worse, even though these phenomic selection models included some of the breeding material in their testing sets as a reference. With phenomic prediction (Figure 2), the SERAT test was predicted relatively well ($r = .53$; Figure 2b) although the GEM dataset from 2011 was predicted with poor accuracy ($r = .33$). Additionally, although tests 1AF and 2AF were also predicted rather poorly ($r = .23$ and $.25$) both were far worse with phenomic selection approaches, where correlations often approached 0. Both of these breeding tests represented material from outside of the original GWAS study. Given that the NIRS BLUP and functional regression

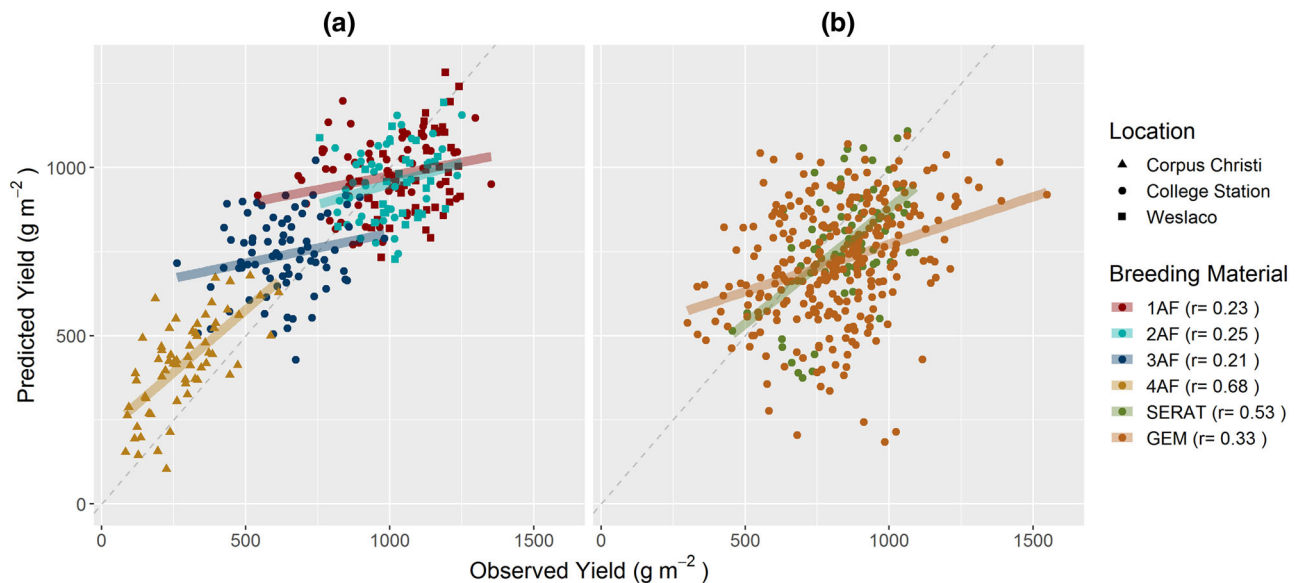


FIGURE 2 Ability for PLSR1 (Table 1) model trained on original samples to predict yield on a plot basis of (a) genetically unrelated samples of breeding trials from the Texas A&M Corn Program in 2011 and (b) genetically unrelated tests representing two tests of material from outside of the program (note: two negatively predicted plots were removed from GEM test). The dotted lines show the 1:1 line

TABLE 8 Repeatability for yield on an entry basis across all environments and within each environment

Environment	Repeatability
ALL	0.7745
WS_2011	0.5559
WW_2011	0.6369
WS_2012	0.7421
WW_2012	0.7534

models could not predict the unrelated genetic material nearly as well as the PLSR models supports that the NIRS BLUP and functional regression rely more on estimating genetic relatedness than the plot-based method.

4.1 | Advantages of NIRS BLUP

The NIRS BLUP model for phenomic selection mimics the genomic best linear unbiased prediction (G-BLUP) model by building a relationship matrix with the bands of the spectral data. This mimicry made NIRS BLUP models very parsimonious despite large p (independent variables, bands in this case) and small n (number of observations). The G-BLUP is one of the most popular methods in genomic selection because of high prediction accuracies of sample data, and because it is easy to process and implement in conventional statistical software. Thus, these benefits are also applicable when computing the NIRS BLUPS via generation of a NIRS similarity matrix between entries.

4.2 | Advantages of functional regression analyses

Functional regression analysis performed very well. Functional regression analyses accounts for spectral bands as transformed covariates of the original bands, which are of considerably fewer dimensions than the original number of bands. Because of this, functional regression is parsimonious and practical to implement in conventional software for phenomic selection. The functional regression analyses was implemented under a Bayesian approach using conventional genomic selection software (Montesinos-Lopez et al., 2018).

One problem with the functional regression model is that it is not easy to choose the required number of basis functions (L) that give rise to the transformed covariates used for final implementation. This process is somewhat analogous to selecting the number PLSR components to use. The correct method for selecting the hyperparameter L is via a tuning process. The advantage is that when the bands (original covariates) are strongly correlated with few transformed covariates ($L = 25$), it must be enough to capture most of the information existing in the original spectral data.

4.3 | Advantages of phenomic selection

Phenomic selection worked well for extracting meaningful information from NIRS of kernels, aligning with findings by Rincent et al. (2018) for predicting end quality traits in wheat. The NIRS BLUP and functional regression both offered parsimonious models for directly including all the

spectral data (bands). The NIRS BLUP included all the spectral data for building a NIRS relationship matrix, which mimics the genomic relationship matrix of the G-BLUP method of genomic selection. The functional regression analyses uses only a few transformed covariates ($L = 101$ in our case) that represent the 3,076 bands originally measured.

An obvious advantage over a simple plot-based prediction method is that the selection methods aim to determine the yield value of a variety, which more easily fits the objectives of most breeders. For breeders it is not necessarily important to understand why the NIRS successfully predicts, simply that the prediction is reliable and consistent within the context that it is used. Cross validation within model development followed by testing on independent tests for robustness can be used to assess reliability.

Genomic prediction has been rapidly adopted by industry in the improvement of major field crops. The genomic selection approach requires building prediction models using genotyping information to better estimate the breeding value of different varieties within a breeding program (Bernardo, 1994; Bernardo & Yu, 2007; Crossa et al., 2010, 2017; Meuwissen, Hayes, & Goddard, 2001; Rincent et al., 2012). Our results suggest that phenomic prediction (such as NIRS of grain yield) is also very useful, although this might be under more restrictive scenarios where the grain composition can be readily assessed, such as in hand-harvested experiments and where NIRS is mounted within grain combines, since NIRS performed best using each sample in each environment. Our results also suggest that phenomic selection, where the NIRS is used to make inferences about the value or performance of a variety, may be even more useful, once protocols for the conditions to best collect NIRS calibration information are better established and understood.

4.4 | Strong potential for NIRS-based phenomic methods

By predicting or at least improving yield estimates, grain NIRS has the potential to reduce the number of field replications needed to compare genotypes. Reducing replications or using less expensive plot evaluation methodology has potential to save breeding resources. Improving phenotypic data accuracy also has potential to improve selection ability. For new technologies to improve plant breeding they need to either reduce evaluation resources, improve data accuracy, or both. High-throughput phenotyping tools like NIRS offer potential for inexpensive, fixed-cost methods of evaluating best yielding varieties within a breeding program. It is important to note that the repeatability for yield in this study was relatively high, which may have impacted the success of NIRS being able to capture yield well across all methods.

Ultimately, high r values, even when the yield predictions are not highly accurate, suggest that NIRS-based models offer a strong potential to be used in ranking varieties relative to one another, as is often done in breeding programs. Furthermore, as suggested by Rincent et al. (2018), there is potential for NIRS to assist in negative selection, to eliminate varieties with confidence that none of the best lines are being lost, before subjecting a population to more expensive or rigorous screening. The most practical future application would be an eventual relative ranking model to be designed to assist breeders in selection.

4.5 | Improving future NIRS phenomics work

Phenomic prediction and selection have a number of potential direct applications. First, it would be useful in improving plot combine yield data, which would be particularly easy to incorporate into programs with NIRS-capable combines. There is also the potential to carry out single-ear, hand-harvested yield estimates to reduce the amount of plants grown or plots combined. This would be most useful when increasing the environments where limited or unskilled labor might be available for collecting some ears but not the entire plot.

One recurrent question surrounding NIRS prediction methods are the optimization of the environments in which to capture scans. Spectra collection in environments separate from the environment in which calibrations are made has presented challenges in the past (Ferrio et al., 2004), and remains a key question to be answered in optimizing selection models using NIRS. It has been shown that NIRS can be used to predict in new environments (Rincent et al., 2018), but understanding what makes the best calibration site is yet to be determined. In this study, where entire environments were held out from the training of the model (CV0), it was always better to build the model using the more stressed environment (e.g., WS_2011), consistent with the findings of Rincent et al. (2018). Surprisingly, the best environment to build the model, the most stressed environment, WS_2011, also had the lowest repeatability. High-stress environments that increase phenotypic differentiation, even if decreasing repeatability, should therefore be used for training.

Beyond single high stress environments, it is likely that the most robust models in a breeding program will be those calibrated with data from multiple years, across a wide diversity of genotypes. This is supported by the improvement made in prediction accuracy between PLSR 5 and PLSR 6 (Table 2), where more genetic material across more environments added to the training set improved prediction results.

One of the expected pitfalls that will vary by crop, is that maize is an outcrossing crop grown as a hybrid, meaning the grain likely is not a pure representation of the mother plant, unlike the successful Rincent et al. (2018) wheat study.

Fortunately, xenia effects are not usually impactful in maize grain (Seka & Cross, 1995). Future work could consider NIRS collection on the hybrid seed before it is planted, but lacking a xenia effect, this would likely only represent the maternal parent and not the hybrid per se.

In terms of improving the statistical models used, using a block diagonal covariance matrix, with an environment-specific \mathbf{W} on the diagonal may be a way to improve the NIRS BLUP models. The fact that this NIRS BLUP model performed better under CV1 than CV2 suggests there may be strong structuring in the \mathbf{W}_A covariance matrix used in this work.

4.6 | Applications to other high-throughput phenotyping technologies

The overall concepts of phenotyping have evolved from the direct measurement of the trait of interest, to investigations surrounding more indirect measures of overall plant growth or spectral signatures, which may be inferred via remote or proximal sensing (Cabrera-Bosquet, Crossa, von Zitzewitz, Serret, & Araus, 2012). Phenotyping concepts are now further expanding to phenomic selection methodologies to estimate genetic relationships. Beyond NIRS, one of the best potential data collection systems to collect the volume of data needed are unoccupied aerial systems (UAS, i.e., drones). Recent reductions in UAS costs and technological hurdles have increased opportunities for new types of data collection by researchers and farmers (Araus & Cairns, 2014; Shi et al., 2016). When UAS campaigns successfully collect images over breeding programs, especially temporally over a growing season, many features from the images can be extracted. Some of these features may correspond with traditionally measured phenotypes such as plant height (Pugh et al., 2018), which themselves can be incorporated into physiological growth models to extract novel, more robust phenotypes (Anderson et al., 2019). Researchers have excitingly begun to incorporate many features into statistical models that can predict end yield, such as plant height (Anderson et al., 2019) and vegetation indices (Sun et al., 2019), already known to have some relationship with yield. As the numbers of phenotypic features are expanded these may complement, or replace, current genomics methods. In wheat, developing relationship matrices from hyperspectral data collected via UAS generated successful yield predictions alone, and using them alongside genetic marker and pedigree information improved predictions in some cases (Krause et al., 2019). Evaluating the thousands of features provided by NIRS can serve as a guide to utilizing many kinds of dense phenotypic datasets. In future work, it is likely that hundreds or thousands of features from UAS temporal flights and hyperspectral sensors could be identified that consistently differ-

entiate between cultivars in the field and applied in phenomic selection the way NIRS spectral bands were here. Most such features have not yet been evaluated for their ability to predict end yield (Shakoor, Northrup, Murray, & Mockler, 2019).

5 | CONCLUSIONS

This study supports previous evidence from wheat that the use of grain spectral data is useful for predicting grain yield, but for the first time, shows that this approach is valid in maize. For this reason, these results offer promise for studies in other crops, and will likely guide users to choose a statistical analysis method that best suits their goals. Partial least squares regression is routinely used with NIRS and showed promise for predicting a non-compositional trait on a plot basis, including on unrelated material. However, functional regression generally performed better, within a set of similar germplasm, and may offer even more value for breeders once protocols are established.

There is reason to suggest that building NIRS prediction models including multiple years will further strengthen the prediction accuracy. Future experiments or applications of these methods should aim for more than 2 yr of data to predict subsequent years, especially focused on higher-stress environments. Low cost of implementation and evidence provided in this work suggest that breeders interested in this technology should scan as much relevant genetic material as possible when building calibrations, and continually add and update the models just as in genomic selection.

ACKNOWLEDGMENTS

The authors would like to acknowledge the students of the Texas A&M Quantitative Genetics and Maize Breeding Program who assisted in making the phenotypic data collection possible. We would also like to thank Dr. Sofija Bozinovic, who tested some of the R scripts for this article. We also acknowledge the reviewers of this manuscript, whose feedback was very valuable to significantly improving the quality of this article.

AUTHOR CONTRIBUTIONS

H. M. Lane: conceptualization, formal analysis, investigation, methodology, software, writing (original draft), review and editing (lead), visualization; S. C. Murray: conceptualization, funding acquisition, methodology, project administration, supervision, resources, review and editing (supporting); O. A. Montesinos-López: software, formal analysis, writing, review and editing (supporting); A. Montesinos-López: software, formal analysis, writing; J. Crossa: software, formal analysis, writing, resources, supervision, review and editing (supporting); D. K. Rooney: data curation; I. D. Barrero-Farfan: data

curation; G. N. De La Fuente: data curation; C. L. S. Morgan: methodology, review and editing (supporting).

FUNDING

This project was made possible by financial support from USDA-NIFA-AFRI Awards 2010-85117-20539, 2017-67013-26185, USDA-NIFA Hatch funds, Texas A&M AgriLife Research, the Texas Corn Producers Board, and the Eugene Butler Endowed Chair. Holly M. Lane was funded under Texas A&M College of Agriculture and Life Sciences Excellence Fellowship.


CONFLICT OF INTEREST

The authors declare there are no conflicts of interest.

ORCID

Holly M. Lane  <https://orcid.org/0000-0001-7047-8905>

Seth C. Murray  <https://orcid.org/0000-0002-2960-8226>


Osva A. Montesinos-López 

<https://orcid.org/0000-0003-4464-3385>

José Crossa  <https://orcid.org/0000-0001-9429-5855>

Gerald N. De La Fuente 

<https://orcid.org/0000-0002-3450-5053>

Cristine L. S. Morgan 

<https://orcid.org/0000-0001-9836-0669>

REFERENCES

- Anderson, S. L., Mahan, A. L., Murray, S. C., & Klein, P. E. (2018). Four parent maize (FPM) population: Effects of mating designs on linkage disequilibrium and mapping quantitative traits. *The Plant Genome*, *11*(2), 170102. <https://doi.org/10.3835/plantgenome2017.11.0102>
- Anderson, S. L., Murray, S. C., Malambo, L., Ratcliff, C., Popescu, S., Cope, D., ... 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*:190004. <https://doi.org/10.2135/tpj2019.02.0004>
- 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>
- Barrero-Farfán, I. D. B., Fuente, G. N. D. L., Murray, S. C., Isakeit, T., Huang, P.-C., Warburton, M., ... Kolomiets, M. (2015). Genome wide association study for drought, aflatoxin resistance, and important agronomic traits of maize hybrids in the sub-tropics. *PLoS ONE*, *10*(2), e0117737. <https://doi.org/10.1371/journal.pone.0117737>
- Bernardo, R. (1994). Prediction of maize single-cross performance using RFLPS and information from related hybrids. *Crop Science*, *34*(1), 20–25. <https://doi.org/10.2135/cropsci1994.0011183X003400010003x>
- Bernardo, R. (2008). Molecular markers and selection for complex traits in plants: Learning from the last 20 years. *Crop Science*, *48*(5), 1649–1664. <https://doi.org/10.2135/cropsci2008.03.0131>
- Bernardo, R., & Yu, J. (2007). Prospects for genomewide selection for quantitative traits in maize. *Crop Science*, *47*(3), 1082–1090. <https://doi.org/10.2135/cropsci2006.11.0690>
- Bertrand, D., Robert, P., & Loisel, W. (1985). Identification of some wheat varieties by near infrared reflectance spectroscopy. *Journal of the Science of Food and Agriculture*, *36*(11), 1120–1124. <https://doi.org/10.1002/jsfa.2740361114>
- Brummer, E. C., Barber, W. T., Collier, S. M., Cox, T. S., Johnson, R., Murray, S. C., ... Thro, A. M. (2011). Plant breeding for harmony between agriculture and the environment. *Frontiers in Ecology and the Environment*, *9*(10), 561–568. <https://doi.org/10.1890/100225>
- Cabrera-Bosquet, L., Crossa, J., von Zitzewitz, J., Serret, M. D., & Araus, J. L. (2012). High-throughput phenotyping and genomic selection: The frontiers of crop breeding converge. *Journal of Integrative Plant Biology*, *54*(5), 312–320. <https://doi.org/10.1111/j.1744-7909.2012.01116.x>
- Chen, J., Ren, X., Zhang, Q., Diao, X., & Shen, Q. (2013). Determination of protein, total carbohydrates and crude fat contents of foxtail millet using effective wavelengths in NIR spectroscopy. *Journal of Cereal Science*, *58*(2), 241–247. <https://doi.org/10.1016/j.jcs.2013.07.002>
- Christman, J. L. (2017). *Development of near infrared reflectance spectroscopy (NIRS) calibrations of maize kernel phosphorus for the identification of useful breeding material*. (Master's thesis, Texas A&M University). Retrieved from <https://oaktrust.library.tamu.edu/handle/1969.1/161508>
- Cooper, M., Smith, O. S., Graham, G., Lane, A., Feng, L., & Podlich, D. W. (2004). Genomics, genetics, and plant breeding: A private sector perspective. *Crop Science*, *44*(6), 1907–1913. <https://doi.org/10.2135/cropsci2004.1907>
- Crossa, J., de los Campos, G., Pérez-Rodríguez, P., Gianola, D., Burgueño, J., Araus, J. L., ... Braun, J. H. (2010). Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics*, *186*, 713–724.
- Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O. A., Jarquín, D., de los Campos, G., ... Varshney, R. (2017). Genomic selection in plant breeding: Methods, models, and perspectives. *Trends in Plant Science*, *11*(22), 961–975. <https://doi.org/10.1016/j.tplants.2017.08.011>
- De La Fuente, G. N., Murray, S. C., Isakeit, T., Park, Y.-S., Yan, Y., Warburton, M. L., & Kolomiets, M. V. (2013). Characterization of genetic diversity and linkage disequilibrium of ZmLOX4 and ZmLOX5 loci in maize. *PLoS ONE*, *8*(1), e53973. <https://doi.org/10.1371/journal.pone.0053973>
- Espinoza, J. A., Hodge, G. R., & Dvorak, W. S. (2012). The potential use of near infrared spectroscopy to discriminate between different pine species and their hybrids. *Journal of Near Infrared Spectroscopy*, *20*(4), 437–447. <https://doi.org/10.1255/jnirs.1006>
- Ferreira, D. S., Pallone, J. A. L., & Poppi, R. J. (2013). Fourier transform near-infrared spectroscopy (FT-NIRS) application to estimate Brazilian soybean [*Glycine max* (L.) Merrill] composition. *Food Research International*, *51*(1), 53–58. <https://doi.org/10.1016/j.foodres.2012.09.015>
- Ferrio, J. P., Bertran, E., Nachit, M. M., Català, J., & Araus, J. L. (2004). Estimation of grain yield by near-infrared reflectance spectroscopy in durum wheat. *Euphytica*, *137*(3), 373–380. <https://doi.org/10.1023/B:EUPH.0000040523.52707.1e>
- Flint-Garcia, S. A., Thuillet, A.-C., Yu, J., Pressoir, G., Romero, S. M., Mitchell, S. E., ... Buckler, E. S. (2005). Maize association population: A high-resolution platform for quantitative trait locus dissection. *The Plant Journal*, *44*(6), 1054–1064. <https://doi.org/10.1111/j.1365-313X.2005.02591.x>

- Foley, W. J., McIlwee, A., Lawler, I., Aragones, L., Woolnough, A. P., & Berding, N. (1998). Ecological applications of near infrared reflectance spectroscopy—a tool for rapid, cost-effective prediction of the composition of plant and animal tissues and aspects of animal performance. *Oecologia*, *116*(3), 293–305. <https://doi.org/10.1007/s004420050591>.
- Krause, M. R., González-Pérez, L., Crossa, J., Pérez-Rodríguez, P., Montesinos-López, O., Singh, R. P., ... Mondal, S. (2019). Hyper-spectral reflectance-derived relationship matrices for genomic prediction of grain yield in wheat. *G3: Genes, Genomes, Genetics*, *9*(4), 1231–1247. <https://doi.org/10.1534/g3.118.200856>
- Lang, C., Almeida, D. R. A., & Costa, F. R. C. (2017). Discrimination of taxonomic identity at species, genus and family levels using Fourier Transformed Near-Infrared Spectroscopy (FT-NIR). *Forest Ecology and Management*, *406*, 219–227. <https://doi.org/10.1016/j.foreco.2017.09.003>
- Lee, R. B., & DeVore, I. (1968). Problems in the study of hunters and gatherers. In R. Lee & I. DeVore (Eds.), *Man the hunter* (pp. 3–12). Chicago: Aldine.
- Levy, S. E., & Myers, R. M. (2016). Advancements in next-generation sequencing. *Annual Review of Genomics and Human Genetics*, *17*(1), 95–115. <https://doi.org/10.1146/annurev-genom-083115-022413>
- Lopez-Cruz, M., Crossa, J., Bonnett, D., Dreisigacker, S., Poland, J., Jan-nink, J. L., ... de los Campos, G. (2015). Increased prediction accuracy in wheat breeding trials using a marker × environment interaction genomic selection model. *G3: Genes, Genomes, Genetics*, *5*(4), 569–582. <https://doi.org/10.1534/g3.114.016097>
- Meng, Q., Murray, S. C., Mahan, A., Collison, A., Yang, L., & Awika, J. (2015). Rapid estimation of phenolic content in colored maize by near-infrared reflectance spectroscopy and its use in breeding. *Crop Science*, *55*(5), 2234–2243. <https://doi.org/10.2135/cropsci2014.11.0767>.
- Meuwissen, T. H. E., Hayes, B. J., & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, *157*(4), 1819–1829.
- Mevik, B.-H. & Wehrens, R. (2007). The pls package: Principal component and partial least squares regression in R. *Journal of Statistical Software*, *18*(2), 1–23. <https://doi.org/10.18637/jss.v018.i02>
- Montesinos-López, O. A., Montesinos-López, A., Crossa, J., de los Campos, G., Alvarado, G., Mondal, S., ... González-Pérez, L. (2017a). Predicting grain yield using canopy hyperspectral reflectance in wheat breeding data. *Plant Methods*, *13*(4). <https://doi.org/10.1186/s13007-016-0154-2>
- Montesinos-López, O. A., Montesinos-López, O. A., Cuevas, J., Mata-López, W. A., Burgueño, J., Mondale, S., ... Crossa, J. (2017b). Genomic Bayesian functional regression models with interactions for predicting wheat grain yield using hyper-spectral image data. *Plant Methods*, *13*, 62. <https://doi.org/10.1186/s13007-017-0212-4>
- Montesinos-Lopez, A., Montesinos-Lopez, O. A., de los Campos, G., Crossa, J., Burgueño, J., & Luna-Vazquez, F. J. (2018). Bayesian functional regression as an alternative statistical analysis of high-throughput phenotyping data of modern agriculture. *Plant Methods*, *14*, 46. <https://doi.org/10.1186/s13007-018-0314-7>
- Morris, J. S. (2015). Functional regression. *Annual Review of Statistics and Its Application*, *2*, 321–359. <https://doi.org/10.1146/annurev-statistics-010814-020413>
- Morris, L. J., & Holman, R. T. (1961). Naturally occurring epoxy acids: II. Detection and measurement of long-chain epoxy acids by near infrared spectrophotometry. *Journal of Lipid Research*, *2*(1), 77–82.
- Murray, S. C., Mayfield, K., Pekar, J., Brown, P., Lorenz, A., Isakeit, T., ... Betran, J. (2019). Tx741, Tx777, Tx779, Tx780, and Tx782 inbred maize lines for yield and southern United States stress adaptation. *Journal of Plant Registrations*, *13*(2): 258–269. <https://doi.org/10.3198/jpr2017.07.00444crp>
- Murray, S. C., Rooney, W. L., Mitchell, S. E., Sharma, A., Klein, P. E., Mullet, J. E., & Kresovich, S. (2008a). Genetic improvement of sorghum as a biofuel feedstock: II. QTL for stem and leaf structural carbohydrates. *Crop Science*, *48*(6), 2180–2193. <https://doi.org/10.2135/cropsci2008.01.0068>
- Murray, S. C., Sharma, A., Rooney, W. L., Klein, P. E., Mullet, J. E., Mitchell, S. E., & Kresovich, S. (2008b). Genetic improvement of sorghum as a biofuel feedstock: I. QTL for stem sugar and grain nonstructural carbohydrates. *Crop Science*, *48*(6), 2165–2179. <https://doi.org/10.2135/cropsci2008.01.0016>
- Park, Y. S., Kunze, S., Ni, X., Feussner, I., & Kolomiets, M. V. (2010). Comparative molecular and biochemical characterization of segmentally duplicated 9-lipoxygenase genes ZmLOX4 and ZmLOX5 of maize. *Planta*, *231*(6), 1425–1437.
- Pugh, N. A., Horne, D. W., Murray, S. C., Carvalho, G., Malambo, L., Jung, J., ... Roone, W. L. (2018). Temporal estimates of crop growth in sorghum and maize breeding enabled by unmanned aerial systems. *The Plant Phenome Journal*, *1*, 170006. <https://doi.org/10.2135/tppj2017.08.0006>
- R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rincint, R., Charpentier, J.-P., Faivre-Rampant, P., Paux, E., Gouis, J. L., 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>
- Rincint, R., Laloë, D., Nicolas, S., Altmann, T., Brunel, D., Revilla, P., ... Moreau, L. (2012). Maximizing the reliability of genomic selection by optimizing the calibration set of reference individuals: Comparison of methods in two diverse groups of maize inbreds (*Zea mays* L.). *Genetics*, *192*(2), 715–728. <https://doi.org/10.1534/genetics.112.141473>
- Rinnan, Å., van den Berg, F., & Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends in Analytical Chemistry*, *28*(10), 1201–1222. <https://doi.org/10.1016/j.trac.2009.07.007>
- Romay, M. C., Millard, M. J., Glaubitz, J. C., Peiffer, J. A., Swarts, K. L., Casstevens, T. M., ... Gardner, C. A. (2013). Comprehensive genotyping of the USA national maize inbred seed bank. *Genome Biology*, *14*(6), R55. <https://doi.org/10.1186/gb-2013-14-6-r55>
- Saint Pierre, C., Burgueño, J., Crossa, J., Fuentes Dávila, G., Figueroa López, P., Solís Moya, E., ... Singh, S. (2016). Genomic prediction models for grain yield of spring bread wheat in diverse agro-ecological zones. *Scientific Reports*, *6*, 27312. <https://doi.org/10.1038/srep27312>
- Seka, D., & Cross, H. Z. (1995). Xenia and maternal effects on maize agronomic traits at three plant densities. *Crop Science*, *35*(1), 86–90. <https://doi.org/10.2135/cropsci1995.0011183X003500010015x>
- Shakoor, N., Northrup, D., Murray, S., & Mockler, T. C. (2019). Big data driven agriculture: Big data analytics in plant breeding, genomics, and the use of remote sensing technologies to advance crop productivity. *The Plant Phenome Journal*, *2*(1), 180009. <https://doi.org/10.2135/tppj2018.12.0009>

- Shi, Y., Thomasson J. A., Murray S. C., Pugh N. A., Rooney W. L., Shafian, S., ... Yang, C. (2016). Unmanned aerial vehicles for high-throughput phenotyping and agronomic research. *PLoS ONE* 11(7), e0159781. <https://doi.org/10.1371/journal.pone.0159781>
- Silva, C. F. L., Milach, S. C. K., Silva, S. D. A., & Montero, C. R. (2008). Near infrared reflectance spectroscopy (NIRS) to assess protein and lipid contents in *Avena sativa* L. *Crop Breeding and Applied Biotechnology*, 8(2), 127–133. <https://doi.org/10.12702/1984-7033.v08n02a05>
- Stevens, A., & Ramirez-Lopez, L. (2014). An introduction to the prospectr package. R Package Version 0.1 3.
- Sun, J., Poland, J. A., Mondal, S., Crossa, J., Juliana, P., Singh, R. P., ... Sorrells, M. E. (2019). High-throughput phenotyping platforms enhance genomic selection for wheat grain yield across populations and cycles in early stage. *Theoretical and Applied Genetics*, 132, 1705–1720. <https://doi.org/10.1007/s00122-019-03309-0>
- Wahl, N., Murray, S. C., Isakeit, T., Krakowsky, M., Windham, G. L., Williams, W. P., ... Betran, J. (2017). Identification of resistance to aflatoxin accumulation and yield potential in maize hybrids in the Southeast Regional Aflatoxin Trials (SERAT). *Crop Science*, 57(1), 202–215. <https://doi.org/10.2135/cropsci2016.06.0519>
- Warburton, M. L., Williams, W. P., Windham, G. L., Murray, S. C., Xu, W., Hawkins, L. K., & Duran, J. F. (2013). Phenotypic and genetic characterization of a maize association mapping panel developed for the identification of new sources of resistance to *Aspergillus flavus* and aflatoxin accumulation. *Crop Science*, 53(6), 2374–2383. <https://doi.org/10.2135/cropsci2012.10.0616>
- Wolfrum, E., Payne, C., Stefaniak, T., Rooney, W., Dighe, N., Bean, B., & Dahlberg, J. (2013). *Multivariate calibration models for sorghum composition using near-infrared spectroscopy*. Golden, CO: National Renewable Energy Lab (NREL).
- Zhao, Y., Mette, M. F., & Reif, J. C. (2015). Genomic selection in hybrid breeding. *Plant Breeding*, 134(1), 1–10. <https://doi.org/10.1111/pbr.12231>

SUPPORTING INFORMATION

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

How to cite this article: Lane HM, Murray SC, Montesinos-López OA, et al. Phenomic selection and prediction of maize grain yield from near-infrared reflectance spectroscopy of kernels. *The Plant Phenome J.* 2020;3:e20002. <https://doi.org/10.1002/ppj.2.20002>