

ORIGINAL ARTICLE

Phenomic versus genomic prediction—A comparison of prediction accuracies for grain yield in hard winter wheat lines

Zachary J. Winn¹  | Amanda L. Amsberry¹ | Scott D. Haley¹  | Noah D. DeWitt²  | Richard Esten Mason¹

¹Department of Soil and Crop Sciences, Colorado State University, Fort Collins, Colorado, USA

²School of Plant, Environmental, and Soil Sciences, Louisiana State University, Baton Rouge, Louisiana, USA

Correspondence

Zachary J. Winn and Richard Esten Mason, Department of Soil and Crop Sciences, Colorado State University, Fort Collins, CO 80523, USA. Email: zwinn@outlook.com and esten.mason@colostate.edu

Assigned to Associate Editor Weizhen Liu.

Funding information

WheatCAP, Grant/Award Number: 2022-68013-36439; USDA National Institute of Food and Agriculture

Abstract

Common bread wheat (*Triticum aestivum* L.) is a key component of global diets, but the genetic improvement of wheat is not keeping pace with the growing demands of the world's population. To increase efficiency and reduce costs, breeding programs are rapidly adopting the use of unoccupied aerial vehicles to conduct high-throughput spectral analyses. This study examined the effectiveness of multispectral indices in predicting grain yield compared to genomic prediction. Multispectral data were collected on advanced generation yield nursery trials during the 2019–2021 growing seasons in the Colorado State University Wheat Breeding Program. Genome-wide genotyping was performed on these advanced generations and all plots were harvested to measure grain yield. Two methods were used to predict grain yield: genomic estimated breeding values (GEBVs) generated by a genomic best linear unbiased prediction (gBLUP) model and phenomic phenotypic estimates (PPEs) using only spectral indices via multiple linear regression (MLR), k-nearest neighbors (KNNs), and random forest (RF) models. In cross-validation, PPEs produced by MLR, KNN, and RF models had higher prediction accuracy ($\bar{r} : 0.41 \leq \bar{r} \leq 0.48$) than GEBVs produced by gBLUP ($\bar{r} = 0.35$). In leave-one-year-out forward validation using only multispectral data for 2020 and 2021, PPEs from MLR and KNN models had higher prediction accuracy of grain yield than GEBVs of those same lines. These findings suggest that a limited number of spectra may produce PPEs that are more accurate than or equivalently accurate as GEBVs derived from gBLUP, and this method should be evaluated in earlier development material where sequencing is not feasible.

Abbreviations: AGL, above ground level; AR1, autoregressive order 1; ARDEC, Agricultural Research, Development and Education Center; AYN, advanced yield nursery; CSU, Colorado State University; DH, doubled haploid; DNA, deoxyribonucleic acid; ExGI, excess green index; gBLUP, genomic best linear unbiased prediction; GCP, ground control point; GEBV, genomic estimated breeding value; GS, genomic selection; KNN, k-nearest neighbor; MLR, multiple linear regression; NDVI, normalized difference vegetation index; NIR, near infrared; PPE, phenomic phenotypic estimate; PS, phenomic selection; PYN, preliminary yield nursery; RF, random forest; RGB, red–green–blue; UAV, unoccupied aerial vehicle.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *The Plant Phenome Journal* published by Wiley Periodicals LLC on behalf of American Society of Agronomy and Crop Science Society of America.

1 | INTRODUCTION

Common bread wheat (*Triticum aestivum* L.) is a diet staple across cultures and plays a key role in international food security. The global population is expected to rise to nearly 10 billion people by the year 2050, yet annual genetic gain in wheat yield is not on track to meet this demand, which implies that a biocapacity deficit is inevitable unless significant action is taken (Alexandratos & Bruinsma, 2012; Berners-Lee et al., 2018; Ray et al., 2013). Genomic prediction is the process of using historical genomic and phenotypic data to produce genomic estimated breeding values (GEBVs) for lines that have only been sequenced and have not been evaluated via phenotyping (Meuwissen et al., 2001). Since the advent of high-throughput genomic technologies, breeders have been using these predictions to implement genomic selection (GS) to increase selection intensity and heritability of selection in earlier generation material.

While GS greatly assists breeders in making selections, it is not applied across all generations due to cost and labor limitations. High-throughput spectral analysis via unoccupied aerial vehicles (UAVs) is becoming more common in modern breeding programs, generating large datasets of spectral reflectance and other phenotypic data. While the cost of collecting genome-wide markers on lines becomes prohibitive as the number of lines increases exponentially in earlier generation trials, spectral data may be taken on all planted lines in all fields in a program at a fraction of the cost.

Recent research has suggested the utility of “phenomic selection” (PS) in breeding. Phenomic selection is the process of using electromagnetic spectral reflectance of organisms, captured by multispectral or hyperspectral sensors, to produce phenomic phenotypic estimates (PPEs) for traits of interest (Krause et al., 2019; Lane et al., 2020; Rincent et al., 2018; Robert et al., 2022; Sandhu et al., 2022). Recent publications have illustrated that spectral reflectance can create relationship matrices among individuals; furthermore, it has been shown that spectral reflectance matrices perform at or near the predictive ability of additive relationship matrices derived from either molecular markers or pedigree information when predicting grain yield (Krause et al., 2019). This may imply that PS could stand in place of GS for selection candidates in earlier generations than those lines in the sequencing pipeline.

The potential application of PS in programs will be directly related to access of relevant technologies. Many of the publications on PS either rely on hyperspectral imaging derived from UAVs or near-infrared (NIR) spectrometry taken in the field or in a lab setting (Krause et al., 2019; Lane et al., 2020; Rincent et al., 2018; Sandhu et al., 2022). These methods provide many different bands of reflectance to use for the creation of relationship matrices. However, their use within breeding programs may be limited due to the cost, labor, time, and the processing pipelines required for these techniques. Invest-

Core Ideas

- Phenomic prediction accuracy is comparable to genomic when spectra are heritable and correlated to yield.
- Accurate phenomic phenotypic estimates can be made in unobserved years using prior collected spectral data.
- A limited number of spectral bands collected from multispectral drones can be used in phenomic prediction.

igating if phenomic predictions may be made using a limited number of spectra as predictors on more accessible equipment may address this gap in application.

In the 2019, 2020, and 2021 growing seasons, the Colorado State University (CSU) Wheat Breeding Program flew UAVs in cooperation with the CSU Drone Center (<https://www.research.colostate.edu/csudronecenter/>) over fields containing late generation lines ($F_{3:5}$, $F_{3:6}$, and $F_{3:7+}$) at the Agricultural Research, Development and Education Center (ARDEC, Fort Collins, CO). Concurrently, those late development lines were sequenced for genome-wide markers. Using this unbalanced dataset which models a true breeding system, the objectives of this study were to (1) investigate the use of limited spectral information across years to produce PPEs, (2) compare calculated PPEs and GEBVs to adjusted means of yield, and (3) test several scenarios to observe which models produced optimal results for calculations of PPEs.

2 | MATERIALS AND METHODS

2.1 | Germplasm

The CSU Wheat Breeding Program pipeline is a modified bulk-pedigree method (Figure 1). All lines which occur in the later nurseries of the CSU breeding program are representative of the contemporary hard red and white winter wheat germplasm adapted to the targeted breeding areas of CSU; these areas include eastern Colorado, western Nebraska, western Kansas, and northwestern Oklahoma.

The final single-head selection occurs in an F_3 bulk and selected heads are evaluated in rows in the following year. Final among-row selection occurs in the $F_{3:4}$ and preliminary yield trials begin with $F_{3:5}$ lines. The preliminary yield nursery (PYN; $F_{3:5}$) is then followed by the advanced yield nursery (AYN; $F_{3:6}$). All lines which make it past the PYN and AYN are then placed in the CSU Elite Trial (ELITE; $F_{5:7+}$), where they remain until they are culled or released as a cultivar.

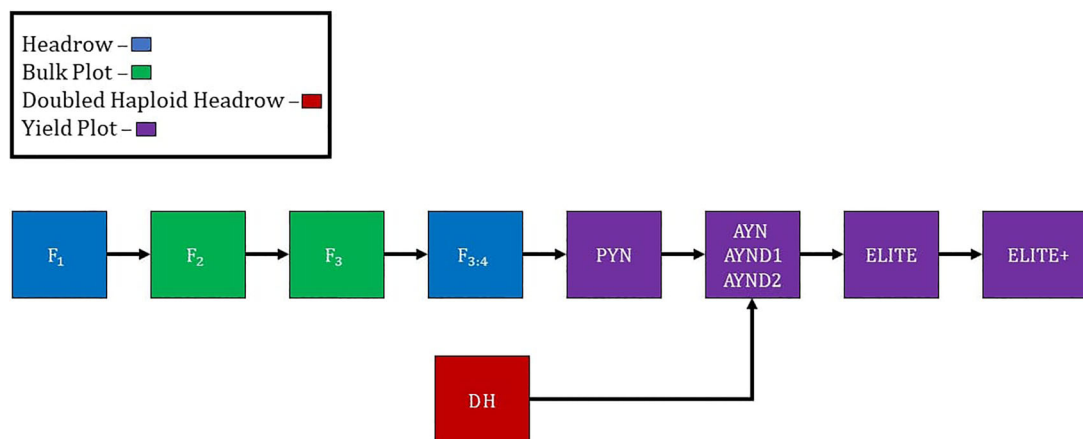


FIGURE 1 A diagram of the Colorado State University (CSU) breeding program. Boxes represent germplasm/trials associated with the first few generations of inbreeding (F_1 to $F_{3:4}$), doubled haploids (DHs), preliminary yield nursery (PYN), advanced yield nursery (AYN), double haploid advanced yield nursery one (AYND1) and two (AYND2), CSU Elite Trial (ELITE), and individuals that stay in the CSU Elite Trial for 3–4 years (ELITE+). Colors represent the plot type of each germplasm/trial. All generations after the $F_{3:4}$ are F_3 -derived lines (e.g., $PYN = F_{3:5}$ and $AYN = F_{3:6}$).

Since 2013, doubled haploids (DHs) have been routinely generated by the CSU Wheat Breeding Program via the F_1 corn (*Zea mays* L.) pollen hybridization method (Santra et al., 2017). After DH seeds have been harvested from greenhouse plants, they are planted in rows and subjected to selection based on visual ratings and genomic predictions. DH lines which are advanced from the row stage are then treated as AYN lines. These nurseries are conventionally titled AYND1 or AYND2 (“D” for DH) within a given year.

Spectral data were collected for 3 years of trials planted at the ARDEC facility in Fort Collins, CO, through collaboration with the CSU Drone Center (<https://www.research.colostate.edu/csudronecenter/>). Data from the following trials were used in the current work: the PYN from 2021; the AYN from 2019, 2020, and 2021; the AYND1 and AYND2 from 2019, 2020, and 2021, and the ELITE from 2019, 2020, and 2021. These trials resulted in a total dataset of 4557 individual observations tied to a yield plot within a trial, within a year. A detailed description of the number of individuals in each nursery for the 2019, 2020, and 2021 season is provided (Table 1).

2.2 | Field management and phenotyping

All nurseries used in this study were planted at the ARDEC facility located in Fort Collins, CO, at approximately 1525 m above sea level. Within each growing season, all grain yield plots were planted in mid-September using a cable guided no-till drill seeder at 1.5-m wide with 4.9-m centers. During green up in the spring, plots were end trimmed using a 1.2-m wide hooded sprayer and glyphosate (Monsanto), leaving a harvestable area of 1.5×3.7 m. All nurseries were planted in

TABLE 1 The number of individual observations and genotypes per generation, per nursery, and year.

Year	Nursery	Number of observations	Number of genotypes
2019	AYN	150	134
	AYND1	598	513
	AYND2	225	193
	ELITE	150	100
2020	AYN	300	242
	AYND1	300	250
	AYND2	400	360
	ELITE	150	100
2021	PYN	1300	1135
	AYN	297	240
	AYND1	297	240
	AYND2	240	120
	ELITE	150	100

Abbreviations: AYN, advanced yield nursery; AYND1, advanced doubled haploid yield nursery 1; AYND2, advanced doubled haploid yield nursery 2; ELITE, CSU Elite Trial; PYN, preliminary yield nursery.

a statistical design which used row–column effects as blocking factors. The PYN, AYN, AYND1, and AYND2 trials were planted in an augmented block design and the ELITE trials were planted in a partially replicated design across all years.

Each year, soil samples were taken across fields to inform pre-plant fertilization regimens. In 2019, 64.61 kg ha^{-1} of 11-52-0 monoammonium phosphate (MAP) granular fertilizer with ammoniacal nitrogen was applied along with $349.83 \text{ kg ha}^{-1}$ 46-0-0 urea granular nitrogen on September 6, 2018, prior to planting. Post planting, 0.95 L ha^{-1} of Husky (BASF)

was applied on April 24, 2019, to control weeds. In 2020, 18.52 kg ha⁻¹ of 0-0-62 high potassium potash was applied along with 32.63 kg ha⁻¹ of 11-52-0 high phosphate MAP granular fertilizer with ammoniacal nitrogen and 358.99 kg ha⁻¹ 46-0-0 urea granular nitrogen on September 13, 2019, prior to planting. Post planting, 1.04 L ha⁻¹ of Husky was applied on April 28, 2020, to control weeds. In 2021, after sampling to determine a proper fertilizing regiment, 32.62 kg ha⁻¹ of 11-52-0 high phosphate MAP granular fertilizer with ammoniacal nitrogen was applied along with 358.18 kg ha⁻¹ 46-0-0 urea granular nitrogen and 15.16 kg ha⁻¹ of granular 90% sulfur on September 6, 2020, prior to planting. Post planting, 1.04 L ha⁻¹ of Husky was applied on May 1, 2021, to control weeds. Irrigation was applied, as needed, as soon as temperatures allowed for irrigation lines to thaw in the spring. Prescriptive irrigation continued into grain fill and was terminated prior to physiological maturity and harvest. Detailed information on irrigation regiments is provided (Table S1).

Within each growing season, at physiological maturity, lines were harvested using a Zürn 150 (Zürn Harvesting GmbH and Company) research plot combine and measured for test weight, grain moisture, and grain yield using a HarvestMaster H2 Classic Graingage (Juniper Systems). Grain yield was adjusted by moisture content so that the grain yields reflected 12% grain moisture, and yield was calculated using plot size metrics and reported in metric tons per hectare on a per-plot basis (t ha⁻¹).

2.3 | Sequencing

For each line sequenced, 10 seeds were planted, and a 2- to 3-cm sample of leaf tissue was taken from each plant and bulked together. Genomic quality deoxyribonucleic acid (DNA) was extracted using MagMax (ThermoFisher Scientific) plant DNA kits. Extracted DNA was quantified using PicoGreen (ThermoFisher Scientific) kits and normalized to a concentration of 20 ng µL⁻¹ using an automated liquid handling system. Sequencing libraries were prepared according to Poland et al. (2012).

All multiplexed libraries were sequenced on a NovaSeq 6000 (Illumina) at 384-plex density per lane. Reads were aligned to the RefSeq Chinese Spring wheat reference sequence v2.0 using the Burrow-Wheeler Aligner version 0.7.17 (Appels et al., 2018; Li & Durbin, 2009). Reads obtained were processed using the TASSEL 2.0 standalone pipeline and single nucleotide polymorphisms (SNPs) were organized into compressed variant calling format files (Danecek et al., 2011; Glaubit et al., 2014).

Variant calling format files were filtered using the following parameters: all SNPs which aligned to an unknown chromosome were removed; SNPs with a read depth of greater than one and less than 100 were retained; all monomorphic

SNPs, insertions, and deletions were removed; all SNPs with a minor allele frequency of 5% or less were removed; all SNPs with a heterozygous frequency of 10% or higher were removed; and all SNPs with a missing data frequency of 85% or higher were removed. To produce full-rank genotypic matrices for genomic relationship matrix calculation, missing data were imputed using the Beagle algorithm V5.4 and the recombination distance-based map used for imputation was derived from a synthetic wheat biparental cross between "W7984" and "Opata" (Browning et al., 2018; Gutierrez-Gonzalez et al., 2019).

2.4 | Spectral reflectance acquisition

The processing pipeline for multispectral data consists of three distinct phases: spectral data collection via UAV-mounted cameras, image stitching and processing, and data extraction from individual plots. Between each year, flight parameters were changed at the discretion of the current drone technician to increase data quality and this resulted in an unbalanced method of sampling from year-to-year. While individual processes and settings may have changed from year-to-year to improve the quality and efficiency of the pipeline over time, the general framework remained similar.

Drone flights were conducted at a height of approximately 120 m above ground level (AGL) in 2019, 100 m AGL in 2020, and 60 m AGL in 2021. This resulted in a resolution of 8.2 px cm⁻¹ in 2019, 7.1 px cm⁻¹ in 2020, and 4.2 px cm⁻¹ in 2021 for the multispectral flights. The increased resolution of the images allowed for more data points per plot when calculating the averages for various spectral indices. In 2019 and 2020, ground control points (GCPs) in the form of plastic bucket lids were placed throughout the field including one at each corner, one at the midpoint of each edge, and at least one in the middle of the field. In 2021, no GCPs were used, which significantly increased the amount of manual post-flight processing.

Flights were conducted at the discretion of the current leader of the CSU Wheat Breeding Program, and flights were taken when scheduling was available through the CSU Drone Center, leading to a wide variance among sampling dates and times. Over several flights taken between 2019 and 2022, the curated flight data have the narrowest window possible. A list of flight dates utilized in this study by year is provided (Table 2).

Two different UAVs were used for flights. The primary UAV was a Matrice 600 Pro (DJI Technology) which housed two cameras—a MicaSense RedEdge-M (AgEagle Aerial Systems) multispectral camera and a FLIR Duo Pro R (Teledyne FLIR) thermal camera. The second UAV was a Phantom 4 RTK (DJI Technology) that gathered a set of backup

TABLE 2 Flight date within year.

Year	Flight 1	Flight 2	Flight 3	Flight 4
2019	4/23/2019	5/24/2019	6/26/2019	7/12/2019
2020	5/1/2020	5/17/2020	6/7/2020	7/2/2020
2021	4/30/2021	5/15/2021	6/8/2021	6/18/2021
Range	8 days	9 days	19 days	24 days

red–green–blue (RGB) data. Between all sensors, data were collected for spectral bands representing blue (465–485 nm), green (550–570 nm), red (663–673 nm), red edge (712–722 nm), NIR (820–860 nm), and long-wave infrared (or thermal; 7.5–13.5 μ m).

All UAVs utilized real time kinematics positioning for accurate geotagging of images. Images were collected in the nadir direction at ± 1 h to solar noon. Immediately before and after each flight, the multispectral camera was calibrated using its accompanying calibration panel. The calibration panel used for each flight was a MicaSense RP04 model panel, which had albedo values of 51.3 for blue at 475 nm, 51.3 for green at 560 nm, 51.3 for red at 668 nm, 51.2 for red edge at 717 nm, and 51 for NIR at 840 nm.

Pix4Dmapper (Pix4D SA) was used to stitch the images together to create an orthomosaic for each of the spectral bands, including the visual bands collected with the RGB camera. These orthomosaics were then uploaded to a geodatabase in ArcGIS Pro (Esri) for further editing and orthorectifying. A single RGB image was selected as the image to georeference all other orthomosaics, which minimizes the geospatial error between orthomosaics and flights over the course of the growing season; this was done by georeferencing each orthomosaic's GCPs to those visible in the reference RGB. In 2021, since no GCPs were placed in the field, other permanent, physical landmarks had to be used for reference.

Starting in 2021, excess green index (ExGI) was used as the source dataset for the ArcGIS machine learning classification function. The goal of using this feature was to remove the potential inaccuracy of the data in previous years by including soil surrounding the plot in the average spectral data calculations. The algorithm was run unsupervised with a high focus on grouping pixels based on both spatial proximity and spectral similarity. The algorithm was tasked with grouping the field into two self-identified categories using the ExGI orthomosaic for reference.

With only plants and soil present in the field, the algorithm successfully classified the field into these two groups. The result of this process was an additional raster layer in the geodatabase that specified boundaries around the plant matter in the field. Using these boundaries, the soil was cropped out of the orthomosaic images, ensuring that the only data remaining in the orthomosaics were related to plant matter. While

retroactive processing of orthomosaics in 2019 and 2020 in this way is possible, the resolution of previous flights was not conducive to separating soil from plant material, thus this procedure was only applied to 2021.

The average reflectance of each band was extracted per plot and ArcGIS Pro was used to calculate spectral indices including normalized difference vegetation index (NDVI). Sampling areas were defined using a Python script that retrieves field coordinates and dimensions from a CSV file and draws a polygon for each plot directly into the ArcGIS geodatabase that hosts the orthomosaics. Using the plot polygons as boundaries for the calculation of the averages, data tables containing the mean reflectance for each plot, date, band, and index were exported for use in further analysis.

2.5 | Validation scenarios

Three separate scenarios were explored to assess how PPEs calculated for each line from plot-level spectral data compared to GEBVs and adjusted mean yield. In all scenarios presented, no genotypes were replicated between the training and test sets to avoid bias in the calculation of GEBVs or PPEs.

Scenario one (S1) involved taking the total available data across years, partitioning 80% of the available yield data to training across years and trials, and validating the remaining 20% while including all spectral data in both training and validation sets. This scenario most closely resembles a classic cross-validation scheme in genomic prediction. This process was replicated 30 times to obtain distributions of cross-validated accuracies.

Scenario two (S2) involved calculating adjusted means on a nursery-by-nursery basis and predicting yield in a trial where yield and spectral data from the same environment are in the training data. In S2, we assume that we have historical spectral and yield information on fields, and we have spectral and yield data on plot-level observations of other nurseries in the same field as a nursery which was not measured for yield. In S2, a single trial is left out of the dataset and predicted using the rest of the available data. This may be applicable in a scenario where harvest of a trial is not feasible and grain yields must instead be predicted. This may be denoted as “hold-one-trial-out forward validation.”

Scenario three (S3) involved taking the total available data within 2 of the 3 available years and using those data to train a model to predict the remaining year. For example, to predict the yield of all genotypes in 2021, both yield and spectral data from 2019 and 2020 and only spectral data from 2021 were used to predict yield data from 2021. This method implies that the program would have historical spectral data on a specific field, and that historical data could be used to train a model to predict yield in an unobserved year

for an unobserved genotype with spectral data. This scenario more closely resembles forward validation in GS and may be denoted as “hold-one-year-out forward validation.”

2.6 | Statistical analysis—Phenotypic analysis and genomic prediction

All data management and statistical analysis was performed in R statistical software version 4.2.2 (R Core Team, 2022). All distributions for spectra and yield were checked for quality and normality using the qqplots function via the Base and Stats packages in R (R Core Team, 2022). All $Q-Q$ plots exhibited near normality, warranting no transformation of any spectral indices or grain yield to fit normality. All mixed linear and genomic prediction models were performed using the package ASReml-R (Butler et al., 2009).

For calculation of adjusted means for both spectral indices and yield, the following univariate mixed linear model was used in every trial:

$$y_{ijk} = \mu + G_i + r_j + c_k + \varepsilon_{ijk}$$

where y_{ijk} is the response, μ is the intercept, G_i is the fixed genotypic response, r_j is the random row effect where $r_j \sim N(0, \sigma_r^2)$, c_k is the random column effect where $c_k \sim N(0, \sigma_c^2)$, and ε_{ijk} is the residual where $\varepsilon \sim N(0, \sigma_e^2)$. To calculate adjusted means across trials within a year or across the total dataset, the following mixed linear model was employed using the adjusted means calculated on a per-trial basis:

$$y_{ij} = \mu + G_i + e_j + \varepsilon_{ij}$$

where y_{ij} is the response, μ is the intercept, G_i is the fixed genotype effect, e_j is the random environment effect where $e_j \sim N(0, \sigma_e^2)$, and ε_{ij} is the residual where $\varepsilon_{ij} \sim N(0, \sigma_e^2)$. An “environment” in this case is the year that the trial was planted, concatenated to the name of the trial, and the location of that trial (e.g., “ARDEC_2019_AYN”). The adjusted means obtained across the entire dataset were used in S1, the adjusted means calculated within each trial were used in S2, and the adjusted means calculated within each year were used in S3.

For calculation of GEBVs in either S1, S2, or S3, the following genomic best linear unbiased prediction (gBLUP) model was employed:

$$y = Xb + Zu + e$$

where y is the response, X is the design matrix for fixed effects, b is the vector of fixed effects, Z is the design matrix for random genetic effects, and u is the vector of additive genetic effects defined by the genomic relationship matrix G ($u \sim$

$N(0, \sigma_G^2)$) (VanRaden, 2008). Relationship matrix G was calculated from markers and weighted by allele frequency.

Genomic, narrow-sense, per-plot, heritabilities (h_g^2) were estimated for each trial in each year using the error and genotypic variances derived from gBLUP model. The gBLUP model used to estimate heritability is the same as presented above; however, no random effects were fit for either row or column effect. Derived variances were then used to calculate h_g^2 as such:

$$h_g^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}$$

where h_g^2 is genomic, narrow sense, per plot heritability; σ_g^2 is genetic variance; and σ_e^2 is the residual variance. This estimate of genomic, narrow sense, per plot heritability is derived from the genetic variance of a gBLUP model divided by the total variance. The genotypic variance was defined by the additive relationship matrix among individuals, and this estimate of heritability is an approximation of narrow-sense heritability, similar to estimates derived by population design when all markers in linkage with causal effects are present in the genetic dataset (de Los Campos et al., 2015). Heritabilities were calculated using the vpredict function in ASReml-R (Butler et al., 2009).

2.7 | Statistical analysis—Phenomic prediction

PPEs were calculated via multiple linear regression (MLR), k-nearest neighbors (KNNs), and random forest (RF). All models used in calculation of PPEs were implemented using the package caret in R (Kuhn, 2008, 2022). In MLR, spectral indices were used as direct regressors on the response (adjusted means for grain yield). The MLR model may be visualized as:

$$y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \dots + \beta_{p-1} x_{i,p-1} + \varepsilon_i$$

where y_i is the response, β_0 is the intercept multiplied by a constant of one, subsequent β_n are the regression coefficients on spectral variables 1 through p , p represents a spectral variable (a spectral index), and the residual error is identically and independently distributed ($\varepsilon_i \sim N(0, \sigma_e^2)$). The estimate derived from the linear model on observations, which only have spectral indices is interpreted as the PPE.

In the case of the nonparametric machine learning algorithms to follow, a seed was set for reproducibility in the case of S2 and S3, and 1000 permutations of fivefold cross-validation in the training set was used to select the optimal hyper-parameters based on lowest achieved residual mean

square error. When KNN models are used for regression of a continuous response, the data in a training set are used to construct a rule on how many nearest neighbors (k) to use in estimating the response in a test set. Estimation of the response is done directly by taking the k number of individuals in the training set most like the unobserved individual and directly averaging them to estimate the unobserved individual (Kramer, 2013; Taunk et al., 2019).

In the KNN algorithm used in this study, tested values of k ranged from 1 to 50 in groups of five (e.g., $k = 1, 5, 10, \dots, 50$). Based on cross-validation hyper-parameter tuning of the training data, we chose the value of k which minimized the residual mean square error. The average among the KNNs used for estimating the unobserved individual in a validation set is interpreted as the PPE in this case.

RF algorithms are based on binary recursive partitioning to generate randomly designed decision trees. RF models can predict categorization by using these randomly generated decision trees to “vote” on categorical responses. However, in RF models used for regression, decision trees cannot take a direct voting approach when the response is continuous. Rather, RF models which are used for regression create cut-off points in the continuous response and use predictor variables (in this case, spectral reflectance data) to define an estimate of the response (grain yield) by averaging observations in the training data set that meet the same criterion.

For continuous predictor variables, finding the best possible split involves organizing the values of the predictors and identifying unique splits between every distinct pair of consecutive values. Normally, the midpoint of the interval between continuous response points is used as the point of division. These splits are then used to bifurcate the tree. In practice, many hundreds of trees are made using this method, and unobserved individuals are passed down these regression trees to form an estimation of the response. The prediction of the continuous response of the unobserved individual is equivalent to the summation of each estimation derived from each random regression tree in the random forest (Biau & Scornet, 2016; Breiman, 2001). This summation is interpreted as the PPE.

For the RF model used in this study, defining variables (spectra) used for randomly splitting nodes in regression trees ranged from 1 to 20 in groups of five (e.g., 1, 5, 10, ..., 20). The number of random decision trees in the RF was optimized using the caret R package. Based on cross-validation hyper-parameter tuning, we selected the n number of random splitting nodes that resulted in lowest residual mean square error in the training set.

To compare the predictive ability of GEBVs against PPEs, the GEBVs and PPEs were regressed on the adjusted means. Genomic and phenomic prediction accuracies were equivalent to the correlation of the corresponding GEBV or PPE to the

adjusted means of grain yield. Both the Pearson's correlation (r) and the coefficient of determination (r^2) were reported as well as the residual mean square error (RMSE). The RMSE of any model mentioned was calculated as such:

$$\text{RMSE} = \sqrt{\frac{\sum (P_i - O_i)^2}{N}}$$

where P_i is the i th prediction, O_i is the i th observations, and N is the total number of observations. In the current work, predictive ability is defined as the Pearson's correlation coefficient between either the GEBVs or PPEs and the adjusted mean relevant to the scenario proposed.

3 | RESULTS

3.1 | Heritabilities and correlations of spectra and grain yield

In general, the genomic, narrow-sense, per-plot heritability of spectra was lower in 2019 and 2020 in comparison to the 2021 growing season (Figure 2). In the 2021 growing season, most of the ranges for estimated heritability of spectra did not include zero in their interval and had smaller ranges than in previous years. The mean, minimum, maximum, and standard deviation of all heritabilities observed within a year for grain yield and all spectra are provided (Table S2).

Correlation of spectral adjusted means to grain yield across trials within each year indicated that spectra were significantly correlated with grain yield in 2021 (Figure 3). However, correlations were, for the most part, sparsely correlated or near zero for the 2019 and 2020 growing seasons. In 2021, most spectral data (RGB, RE, and NIR) collected from later flights had significant negative correlations with grain yield, while NDVI alone had significant positive correlations to yield. The significant positive correlation of grain yield to NDVI implies that plots with higher photosynthetic activity tended to yield higher, a trend consistent with other research in wheat (Freeman et al., 2003; Hassan et al., 2019).

3.2 | Scenario 1: Cross-validation

Randomly partitioned 80:20 training-test cross-validation across 30 permutations revealed that gBLUP models, which only use additive genetic relationship information derived from molecular markers, underperformed in comparison to MLR, KNN, and RF models using only the five spectral indices recorded over four flights (Figure 4a). The average cross-validated prediction accuracy of GEBVs derived from gBLUP was $\bar{r} = 0.35$. By comparison, the average cross-validated prediction accuracies of PPEs derived from the

Average Heritability Within Each Year

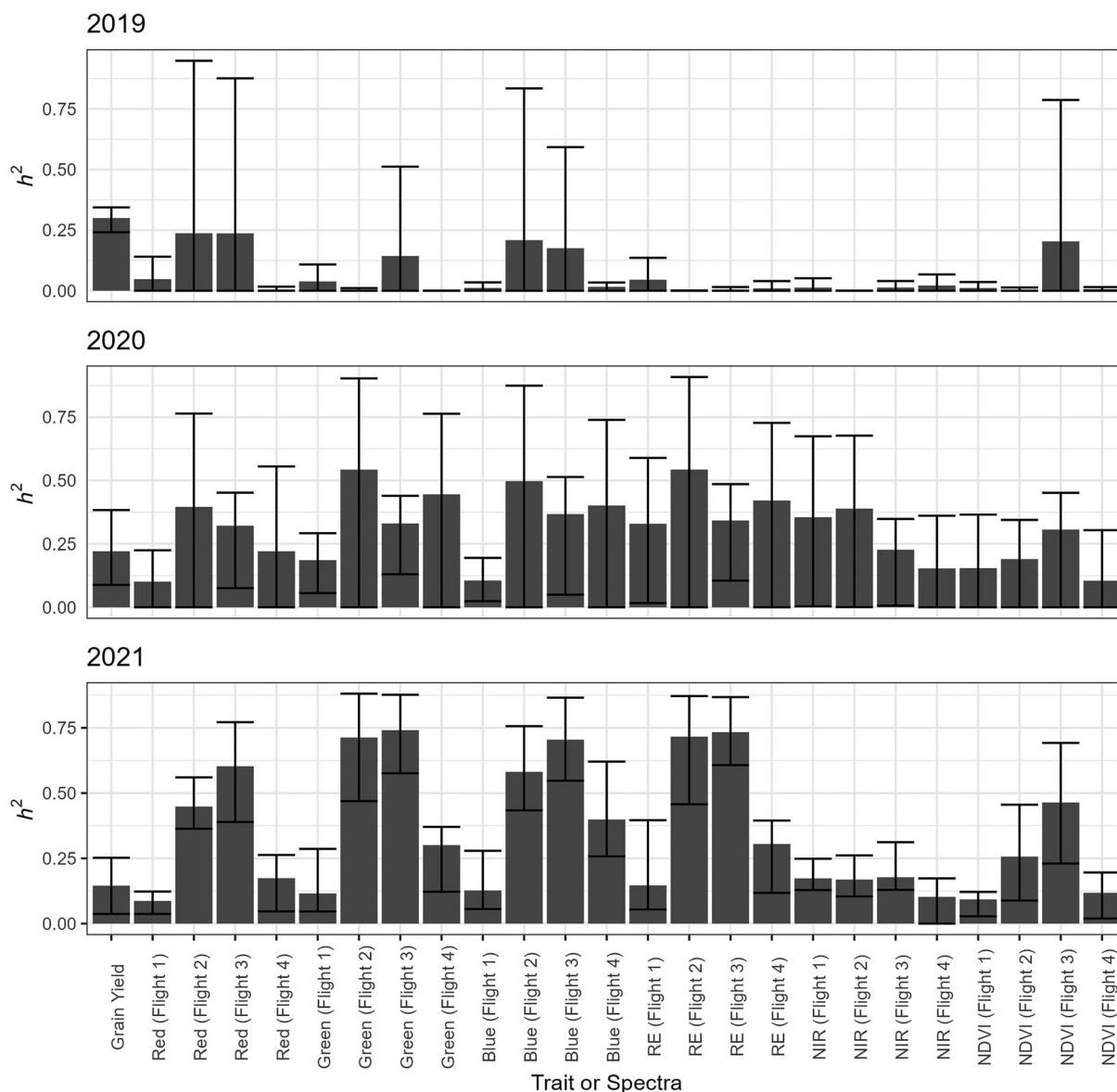


FIGURE 2 Bar-chart of average heritability across trials with each growing cycle. On the y-axis is the narrow-sense, per-plot, and genomic heritability (h^2). The x-axis depicts grain yield, red-green-blue (RGB) spectral reflectance, red edge (RE), near-infrared (NIR), and normalized difference vegetation index (NDVI). The flight from which each spectral index was collected and is noted next to the spectrum name (e.g., Red [Flight 1]). Each bar represents the average heritability of the corresponding trait on the x-axis. The black error lines surrounding each bar represent the observed range for that heritability across nurseries.

MLR, KNN, and RF models were $\bar{r} = 0.41$, $\bar{r} = 0.44$, and $\bar{r} = 0.48$, respectively.

The standard deviations of the gBLUP, MLR, KNN, and RF predictions were $\hat{\sigma}_{\text{gBLUP}} = 0.028$, $\hat{\sigma}_{\text{MLR}} = 0.027$, $\hat{\sigma}_{\text{KNN}} = 0.028$, and $\hat{\sigma}_{\text{RF}} = 0.024$, respectively. A visualization of the prediction accuracy per permutation is provided (Figure 4b). The RMSEs for the gBLUP, MLR, KNN, and RF predictions were $\text{RSM } E_{\text{gBLUP}} = 266.3$, $\text{RSM } E_{\text{MLR}} =$

358.8 , $\text{RSM } E_{\text{KNN}} = 317.9$, and $\text{RSM } E_{\text{RF}} = 332.7$, respectively. None of the models had noticeable differences in the standard deviation of their prediction accuracies, indicating that the range in prediction accuracies across permutations were similar. Of note, there was one prediction accuracy that significantly deviated in permutation seven of the MLR model (Figure 4b). A tabular summary of parameters from cross-validation is also provided (Table S3).

Correlation of Spectra With Grain Yield Within Each Year



FIGURE 3 Bar-chart of Pearson's correlation coefficient (r) of spectra versus grain yield. The x-axis depicts different combinations of spectra and flights. Each bar represents the Pearson's correlation coefficient of the corresponding trait on the x-axis to grain yield. The color of each bar represents the significance of that correlation. The figure legend denoting significance can be found to the right of the graphs. NDVI, normalized difference vegetation index; NIR, near-infrared; NS, not significant; RE, red edge.

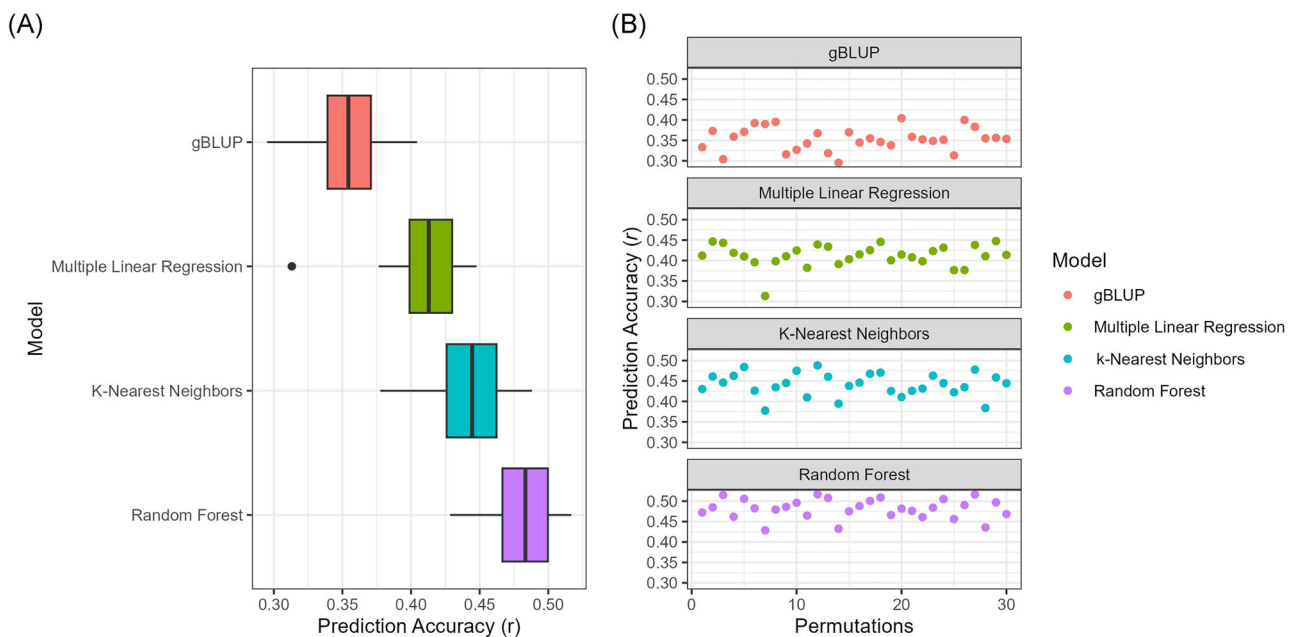


FIGURE 4 Boxplots of cross-validation prediction accuracies (a) and a visualization of prediction accuracy (r) over 30 permutations (b). On the y-axis of panel (a) is the name of the model. On the x-axis of panel (a) is the prediction accuracy. The sub-header of all graphs in panel (b) displays the name of the model. On the y-axis of panel (b) is the prediction accuracy. On the x-axis of panel (b) is the index for each permutation. The figure legend on the far right displays the corresponding colors for each model. gBLUP, genomic best linear unbiased prediction.

3.3 | Scenario 2: Leave-one-trial-out forward validation

In the 2019 season, PPEs underperformed in terms of prediction accuracy (r) in comparison to GEBVs (Figure 5). In the 2020 trials, the only cases where the prediction accuracies for PPEs were comparable to or higher than those derived from GEBVs were in the AYND1 and AYND2. The gBLUP prediction accuracy for the AYND1 in 2020 was $r = 0.33$ as compared to $r = 0.41$ for the MLR model and $r = 0.35$ for the KNN model. Likewise, in the AYND2 in 2020, the prediction accuracy for gBLUP was $r = 0.26$, while the accuracies for MLR and KNN were $r = 0.35$ and $r = 0.29$, respectively.

The trends observed in the 2019 and 2020 data were not apparent in the 2021 data. Overall, PPE prediction accuracies outperformed GEBV prediction accuracies. In the PYN of 2021, the predictive accuracy of the gBLUP model was $r = 0.21$; by comparison, MLR, KNN, and RF had substantially higher prediction accuracies ($r = 0.56$, $r = 0.45$, and $r = 0.57$, respectively). In the AYN, AYND1, AYND2, and ELITE, the prediction accuracies derived from PPE estimates either performed on-par with or better than the prediction accuracy of GEBV in most cases (Figure 5). A table of all model statistics is provided as well (Table S4).

3.4 | Scenario 3: Leave-one-year-out forward validation

Similar to S2, the prediction accuracies of PPEs in 2019 underperformed in comparison to GEBVs ($r_{\text{GEBV}} = 0.15$, $r_{\text{PEBV}} \leq 0.05$; Figure 6). In both 2020 and 2021, PPEs derived from MLR and KNN models consistently outperformed GEBVs derived from gBLUP models. Only in 2021 did the RF model perform above the gBLUP model. In 2020, the following prediction accuracies were observed: $r = 0.18$ for gBLUP, $r = 0.30$ for MLR, $r = 0.25$ for KNN, and $r = 0.01$ for RF. In 2021, the following accuracies were observed: $r = 0.12$ for gBLUP, $r = 0.32$ for MLR, $r = 0.22$ for KNN, and $r = 0.26$ for RF.

4 | DISCUSSION

In the present work, GEBVs were compared to predictions from spectral data (denoted as PPEs) through comparisons of prediction accuracies across several scenarios. In S1, we performed a cross-validation study across 30 permutations and observed that models that produce PPEs using only five spectral indices over four flights were more accurate in terms of correlation with the adjusted mean than GEBVs produced from gBLUP models. In S2, we saw that PPEs were more

accurate than GEBVs when the heritability of the spectral indices was moderate to high and when spectra were significantly correlated with grain yield. In S3, we saw that using data from different years to produce PPEs for an unobserved year produced more accurate results than GEBVs when heritability of spectral indices was moderate to high.

Rincen et al. (2018) and Roberts et al. (2022) suggest that spectral reflectance captures the endophenotypes of organisms and it is this phenomenon that allows for accurate predictions in PS. Furthermore, spectral reflectance can give estimates of plant health, disease pressure, and environmental stress (Du & Noguchi, 2017; Francesconi et al., 2021; Montesinos-López et al., 2017; Qin et al., 2022). This may suggest that PPEs include nonadditive effects like environmental variation in their estimation, which may improve PPE model predictive accuracy.

Krause et al. (2019) illustrated that relationship matrices can be calculated using hyperspectral reflectance and that those relationships derived from those matrices can be just as effective at predicting grain yield as relationship matrices derived from genetic data. Rincen et al. (2018) found that wheat grain yield and heading date could be accurately predicted within and among environments using NIR spectroscopy to produce PPEs. Moreover, Krause et al. (2019) reported that the level of prediction accuracy was highly correlated with the strength of relationship between hyperspectral reflectance and grain yield. Krause et al. (2019) also found that heritability was heterogeneous across timepoints and that this affected prediction accuracy, which is consistent with the current work's results.

Montesinos-López et al. (2017) found that removing "noisy bands" of hyper-spectral reflectance which had heritability lower than $H = 0.50$ increased the predictive accuracy of their models for grain yield in wheat. Furthermore, Montesinos-López et al. (2017) found that prediction accuracies were higher when they used all bands of spectra rather than one or a few vegetative indices. Rincen et al. (2018) illustrated that accurate predictions for poplar tree (Genus *Populus*) bud flesh could also be achieved through phenomic prediction, meaning that this method could have practical application outside wheat. In support of that point, it was found by Lane et al. (2020) that NIR spectroscopy readings from corn can accurately predict grain yield in environments under water stress and well-watered conditions.

In the present work, it was observed that the 2021 data showed higher heritability, higher correlations with grain yield, and improved prediction accuracies compared to those in 2019 and 2020. Two major differences in drone flights and data processing pipelines may be attributed to 2021 being the highest quality data in our dataset. First, the reduction in flight height from 2019 to 2020, and then again in 2020 to 2021, resulted in higher image resolution with each subsequent year, which increased the amount of data points per plot in the

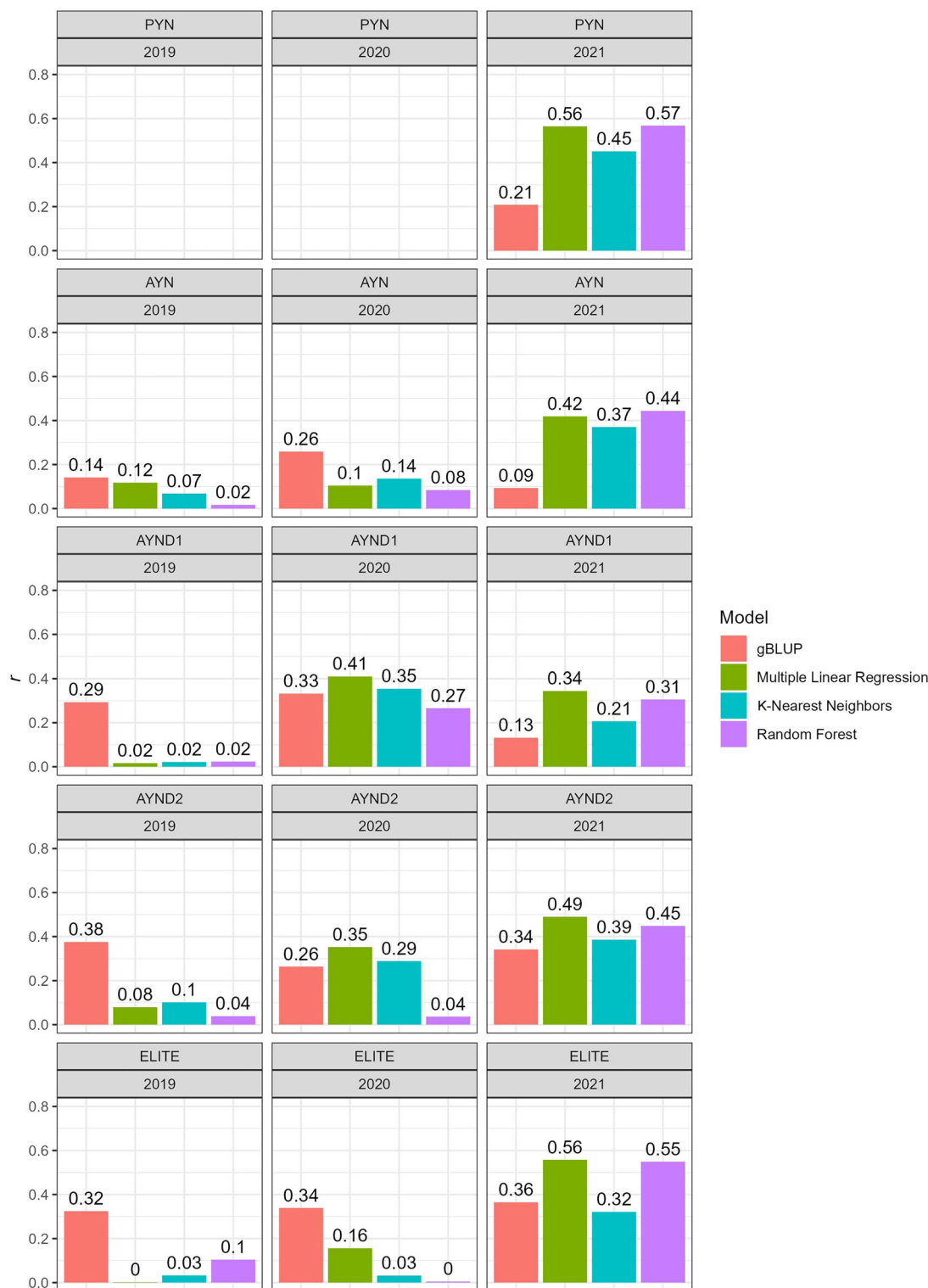


FIGURE 5 Bar graphs of leave-one-trial-out forward validation depicting the prediction accuracy (r) for each trial within each year. The y-axis depicts the prediction accuracy as an expression of the Pearson's correlation coefficient between the estimated breeding value and the best linear unbiased estimate for that trial. Each sub-graph is labeled by the trial to which the contents belong and the year that the trial was planted. Trial abbreviations are as follows: preliminary yield nursery (PYN), advanced yield nursery (AYN), advanced doubled haploid yield nursery one (AYND1), advanced doubled haploid yield nursery two (AYND2), and CSU Elite Trial (ELITE). Bars are colored to denote the model used. The number displayed above each bar represents the prediction accuracy of that model, in that year, for that trial. Subgraphs which do not contain content symbolize the unavailability of that data within that year (e.g., PYN 2019 and 2020). gBLUP, genomic best linear unbiased prediction.

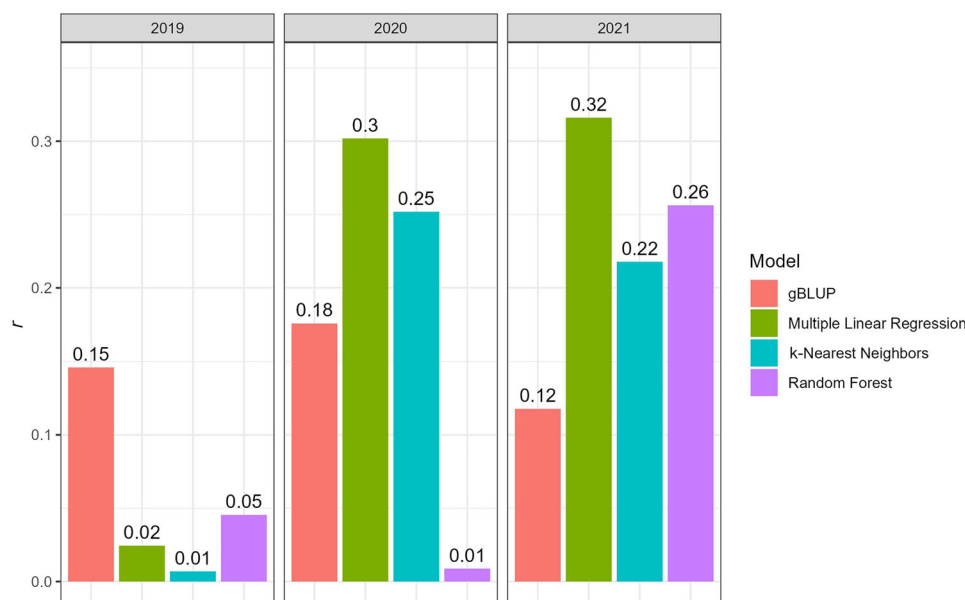


FIGURE 6 Bar graphs of leave-one-year-out forward validation prediction accuracies. The y-axis depicts the prediction accuracy (r). Each subplot has a subtitle above it indicating the year to which the accuracy belongs. Each bar within each graph has been filled with a color that corresponds to what model that bar represents. The figure legend to the right denotes what model is represented by each color. Above each bar, in plain text, is the prediction accuracy. gBLUP, genomic best linear unbiased prediction.

calculation of average reflectance values. This most likely increased the accuracy of spectral averages derived from plots due to an increase in number of data points (pixels) per plot.

The second pipeline improvement which most-likely contributed to 2021 being the most informative data was the implementation of the ExGI to categorize and remove soil from the analysis in 2021. This process implemented in 2021 eliminated the risk of soil within the plot area skewing the averages away from representative plant reflectance values. While this process of classifying and removing soil could have been retroactively performed in previous years for a direct comparison, the flight heights in previous years led to a reduction in resolution, which hinders the programs' ability to discriminate between soil and plant matter accurately. Therefore, we recommend reduction of flight heights to approximately 60 m or less to increase resolution and the use of a similar machine learning method to remove non-plant material from images prior to extraction of spectra from images.

Another contributing factor for varying results across years may be related to flight dates (Table 2). In the current analysis, we took existing drone data taken within the CSU breeding program and aligned flight dates that were closest to each other in time. These data, while apparently informative, are not balanced. Moreover, growth stage in wheat is known to affect spectral reflectance (Sun et al., 2010; Xie et al., 2020). Therefore, to reduce variability in the dataset and potentially improve reproducibility of results, it is suggested to assess fields at the same growth stage across years.

In S1, where we compared 80:20 split cross-validation of GEBVs versus PPEs, we observed that all methods of estimation of PPEs were superior in their prediction accuracies in comparison to GEBVs. Thus, it appears that PPEs have similar predictive ability to GEBVs across years. However, it is important to note that the GEBVs estimated by our gBLUP models are based on additive effects, whereas the PPEs most-likely include environmental and nonadditive effects in their calculation. Therefore, it would be important to conduct follow-up studies on response to selection for PPEs versus GEBVs in early generation material.

While S1 demonstrates the predictability of the dataset upon itself, S2, which involved predicting yield for plots where harvest is not possible, may have some direct utility in breeding programs. These trials in S2, which would be otherwise abandoned due to harvest conditions or labor constraints, may be flown with drones and predictions can be made without genetic information. This information could be used for making selections in limited datasets as well as supplementing training population data in genomic prediction by increasing the number of total observations for the phenotype of interest, yet these applications must be vetted in further studies.

Although S1 and S2 yielded fascinating results, the scenario with the highest utility is S3, where we attempted to predict observations in an unobserved year with historical training data. While it appears that performing this procedure in years where spectral data collection is suboptimal (e.g., spectral data collected in 2019) leads to underperform-

mance in comparison to GEBVs (Figure 6), we did observe that in following years spectral data were more heritable and better correlated to yield, that PPEs were comparable to or better than GEBVs. This may indicate that PPEs may be made on unobserved genotypes in unobserved years so long as the criteria of correlation and heritability are met.

A major component of the cost and labor of phenotyping trials for yield data involves the planting and harvesting of plots. Calculation of PPEs for S3 may allow breeders to plant larger trials of early-generation material and selectively harvest a subset of lines based on collection of spectral data. Although it is possible to use spectral data from plots prior to harvest to produce PPEs, this method only saves effort by allowing for selective harvest, and the same amount of financial and work resources would still need to be invested in planting. In contrast, GS offers a way to reduce planting by using genotyping out of field and prediction, which can be more cost effective over time.

Selective harvest of plots can be challenging with mechanized harvest methods, and this may be less than optimal for programs which do not selectively harvest preliminary yield trials. In lieu of applying this method in the yield plot stage, perhaps the method proposed in S3 would be best applied in the earlier generations where the number of individuals to genotype would be far too large for GS and plots are harvested by hand.

Hypothetically, single plants or headrows of selection candidates could be planted with a subset of lines (a training population), which have replicated yield observations over years. The spectral information and historical yield data of the training population could then be used to produce PPEs for single plants or headrows which only have spectral data. This method is similar to the one proposed by Roberts et al. (2022). Regardless, this method has yet to be tested for within-family prediction/selection, and the response to PS in comparison to visual or GS has not been reported in the literature.

The results of the current work, in correspondence with previously written literature, support the hypothesis that PPEs can be produced from spectra and that PPE accuracies are comparable to or better than GEBV accuracies, if the spectra are heritable and correlated with the trait of interest. More interestingly, we have demonstrated that a limited number of spectral bands, rather than the wide range of spectra derived from hyperspectral analysis, may be used to create comparable predictions.

The presented results indicate that the 2019 spectral data were suboptimal due to sampling methods; however, models in S3, which included the 2019 and 2020 spectral data to predict grain yield in 2021, managed to produce prediction accuracies higher than that of gBLUP (Figure 6), which seems to suggest that large datasets over time can compensate for suboptimal years of data collection. Thus, with further investigation and restructuring of breeding methods, phe-

nomic prediction may be an effective alternative to genomic prediction in earlier generation material.

AUTHOR CONTRIBUTIONS

Zachary J. Winn: Conceptualization; data curation; formal analysis; investigation; methodology; resources; software; supervision; validation; visualization; writing—original draft; writing—review and editing **Amanda L. Amsberry:** Data curation; conceptualization; data curation; methodology; software; supervision; validation; visualization; writing—review and editing. **Scott Haley:** Conceptualization; data curation; resources; writing—review and editing. **Noah DeWitt:** Writing—review and editing. **Richard Esten Mason:** Data curation; funding acquisition; project administration; supervision; writing—review and editing.

ACKNOWLEDGMENTS

This research was made possible by funds derived from the competitive grant 2022-68013-36439 (WheatCAP) from the USDA National Institute of Food and Agriculture.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Relevant code and data used in the presented work can be found at the public GitHub repository (<https://github.com/zjwinn/Phenomic-vs-GP-A-Comparison-of-Prediction-Accuracies-for-GY-in-HWW>).

ORCID

Zachary J. Winn  <https://orcid.org/0000-0003-1543-1527>

Scott D. Haley  <https://orcid.org/0000-0002-1996-9570>

Noah D. DeWitt  <https://orcid.org/0000-0001-9055-993X>

REFERENCES

- Alexandratos, N., & Bruinsma, J. (2012). *World agriculture towards 2030/2050: The 2012 revision*. ESA Working Paper No. 12-03. FAO Agricultural Development Economics Division.
- Appels, R., Eversole, K., Feuillet, C., Keller, B., Rogers, J., Stein, N., Pozniak, C. J., Choulet, F., Distelfeld, A., & Poland, J. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science*, 361(6403), eaar7191.
- Berners-Lee, M., Kennelly, C., Watson, R., & Hewitt, C. (2018). Current global food production is sufficient to meet human nutritional needs in 2050 provided there is radical societal adaptation. *Elementa: Science of the Anthropocene*, 6(1). <https://doi.org/10.1525/elementa.310>
- Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test*, 25(2), 197–227. <https://doi.org/10.1007/s11749-016-0481-7>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Browning, B. L., Zhou, Y., & Browning, S. R. (2018). A one-penny imputed genome from next-generation reference panels. *The American Journal of Human Genetics*, 103(3), 338–348. <https://doi.org/10.1016/j.ajhg.2018.07.015>

- Butler, D., Cullis, B. R., Gilmour, A., & Gogel, B. (2009). *ASReml-R reference manual*. The State of Queensland, Department of Primary Industries and Fisheries.
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCF tools. *Bioinformatics*, 27(15), 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- de Los Campos, G., Sorensen, D., & Gianola, D. (2015). Genomic heritability: What is it? *PLoS Genetics*, 11(5), e1005048. <https://doi.org/10.1371/journal.pgen.1005048>
- Du, M., & Noguchi, N. (2017). Monitoring of wheat growth status and mapping of wheat yield's within-field spatial variations using color images acquired from UAV-camera system. *Remote Sensing*, 9(3), 289. <https://doi.org/10.3390/rs9030289>
- Francesconi, S., Harfouche, A., Maesano, M., & Balestra, G. M. (2021). UAV-based thermal, RGB imaging and gene expression analysis allowed detection of Fusarium head blight and gave new insights into the physiological responses to the disease in durum wheat. *Frontiers in Plant Science*, 12. <https://doi.org/10.3389/fpls.2021.628575>
- Freeman, K. W., Raun, W. R., Johnson, G. V., Mullen, R. W., Stone, M. L., & Solie, J. B. (2003). Late-season prediction of wheat grain yield and grain protein. *Communications in Soil Science and Plant Analysis*, 34(13–14), 1837–1852. <https://doi.org/10.1081/CSS-120023219>
- Glaubit, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q., & Buckler, E. S. (2014). TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. *PLoS One*, 9(2), e90346. <https://doi.org/10.1371/journal.pone.0090346>
- Gutierrez-Gonzalez, J. J., Mascher, M., Poland, J., & Muehlbauer, G. J. (2019). Dense genotyping-by-sequencing linkage maps of two Synthetic W7984 × Opata reference populations provide insights into wheat structural diversity. *Scientific Reports*, 9(1), Article 1793. <https://doi.org/10.1038/s41598-018-38111-3>
- Hassan, M. A., Yang, M., Rasheed, A., Yang, G., Reynolds, M., Xia, X., Xiao, Y., & He, Z. (2019). A rapid monitoring of NDVI across the wheat growth cycle for grain yield prediction using a multi-spectral UAV platform. *Plant Science*, 282, 95–103. <https://doi.org/10.1016/j.plantsci.2018.10.022>
- Kramer, O. (2013). K-nearest neighbors. In O. Kramer (Ed.), *Dimensionality reduction with unsupervised nearest neighbors* (Vol. 51, pp. 13–23). Springer. https://doi.org/10.1007/978-3-642-38652-7_2
- Krause, M. R., González-Pérez, L., Crossa, J., Pérez-Rodríguez, P., Montesinos-López, O., Singh, R. P., Dreisigacker, S., Poland, J., Rutkoski, J., Sorrells, M., Gore, M. A., & Mondal, S. (2019). Hyperspectral 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>
- Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, 28(1), 1–26.
- Kuhn, M. (2022). *caret: Classification and regression training*. <https://CRAN.R-project.org/package=caret>
- Lane, H. M., Murray, S. C., Montesinos-López, O. A., Montesinos-López, A., Crossa, J., Rooney, D. K., Barrero-Farfan, I. D., De La Fuente, G. N., & Morgan, C. L. S. (2020). Phenomic selection and prediction of maize grain yield from near-infrared reflectance spectroscopy of kernels. *The Plant Phenome Journal*, 3(1), e20002. <https://doi.org/10.1002/ppj2.20002>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Meuwissen, T. H., Hayes, B. J., & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157(4), 1819–1829. <https://doi.org/10.1093/genetics/157.4.1819>
- Montesinos-López, O. A., Montesinos-López, A., Crossa, J., de Los Campos, G., Alvarado, G., Suchismita, M., Rutkoski, J., González-Pérez, L., & Burgueño, J. (2017). Predicting grain yield using canopy hyperspectral reflectance in wheat breeding data. *Plant Methods*, 13, Article 4. <https://doi.org/10.1186/s13007-016-0154-2>
- Qin, W., Wang, J., Ma, L., Wang, F., Hu, N., Yang, X., Xiao, Y., Zhang, Y., Sun, Z., Wang, Z., & Yu, K. (2022). UAV-based multi-temporal thermal imaging to evaluate wheat drought resistance in different deficit irrigation regimes. *Remote Sensing*, 14(21), 5608. <https://doi.org/10.3390/rs14215608>
- Poland, J. A., Brown, P. J., Sorrells, M. E., & Jannink, J. L. (2012). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLOS ONE*, 7(2), e32253. <https://doi.org/10.1371/journal.pone.0032253>
- R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Ray, D. K., Mueller, N. D., West, P. C., & Foley, J. A. (2013). Yield trends are insufficient to double global crop production by 2050. *PLoS One*, 8(6), e66428. <https://doi.org/10.1371/journal.pone.0066428>
- Rincint, R., Charpentier, J.-P., Faivre-Rampant, P., Paux, E., Le Gouis, J., Bastien, C., & Segura, V. (2018). Phenomic selection is a low-cost and high-throughput method based on indirect predictions: Proof of concept on wheat and poplar. *G3: Genes, Genomes, Genetics*, 8(12), 3961–3972. <https://doi.org/10.1534/g3.118.200760>
- Robert, P., Brault, C., Rincint, R., & Segura, V. (2022). Phenomic selection: A new and efficient alternative to genomic selection genomic selection (GS). In N. Ahmadi & J. Bartholome (Eds.), *Genomic prediction of complex traits: Methods and protocols* (pp. 397–420). Springer.
- Sandhu, K. S., Merrick, L. F., Sankaran, S., Zhang, Z., & Carter, A. H. (2022). Prospectus of genomic selection and phenomics in cereal, legume and oilseed breeding programs. *Frontiers in Genetics*, 12, 829131. <https://doi.org/10.3389/fgene.2021.829131>
- Santra, M., Wang, H., Seifert, S., & Haley, S. (2017). Doubled haploid laboratory protocol for wheat using wheat-maize wide hybridization. In P. L. Bhalla & M. B. Singh (Eds.), *Wheat biotechnology: Methods and protocols* (pp. 235–249). Humana. https://doi.org/10.1007/978-1-4939-7337-8_14
- Sun, H., Li, M. Z., Zhao, Y., Zhang, Y. E., Wang, X. M., & Li, X. H. (2010). The spectral characteristics and chlorophyll content at winter wheat growth stages. *Guang Pu Xue Yu Guang Pu Fen Xi*, 30(1), 192–196.
- Taunk, K., De, S., Verma, S., & Swetapadma, A. (2019). A brief review of nearest neighbor algorithm for learning and classification. In *2019 International Conference on Intelligent Computing and Control Systems (ICCS)* (pp. 1255–1260). IEEE.
- VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, 91(11), 4414–4423. <http://doi.org/10.3168/jds.2007-0980>
- Xie, Y., Wang, C., Yang, W., Feng, M., Qiao, X., & Song, J. (2020). Canopy hyperspectral characteristics and yield estimation of winter

wheat (*Triticum aestivum*) under low temperature injury. *Scientific Reports*, 10, 244. <https://doi.org/10.1038/s41598-019-57100-8>

SUPPORTING INFORMATION

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

How to cite this article: Winn, Z. J., Amsberry, A. L., Haley, S. D., DeWitt, N. D., & Mason, R. E. (2023). Phenomic versus genomic prediction—A comparison of prediction accuracies for grain yield in hard winter wheat lines. *The Plant Phenome Journal*, 6, e20084. <https://doi.org/10.1002/ppj2.20084>