INTEGRATING HIGH-THROUGHPUT PHENOTYPING, GENOMIC SELECTION, AND SPATIAL ANALYSIS FOR PLANT BREEDING AND MANAGEMENT

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Margaret Rose Krause

August 2019

INTEGRATING HIGH-THROUGHPUT PHENOTYPING, GENOMIC SELECTION, AND SPATIAL ANALYSIS FOR PLANT BREEDING AND MANAGEMENT

Margaret Rose Krause, Ph. D.

Cornell University 2019

Recent advances in high-throughput phenotyping, genomics, and precision agriculture have provided plant breeders and farmers with a wealth of information on the growth and development of crop plants. Methods for effectively leveraging these data resources are needed in order to drive genetic gain in breeding programs and to increase efficiency in farming systems. Three novel approaches for the development and management of high yielding, adapted crop varieties are presented. First, aerial hyperspectral reflectance phenotypes of bread wheat (*Triticum aestivum* L.) were used to develop relationship matrices for the prediction of grain yield within and across environments with genomic selection. Multi-kernel models combining marker/pedigree information with hyperspectral reflectance phenotypes gave the highest accuracies overall; however, improvements in accuracy over single-kernel marker- and pedigree-based models were reduced when correcting for days to heading. Second, aerial phenotypes collected on small, unreplicated plots representing the seed limited stage of wheat breeding programs were evaluated for their potential use as selection criteria for improving grain yield. The aerial phenotypes were shown to be heritable and positively correlated with grain yield measurements evaluated in replicated yield trials. Results also suggest that selection on aerial phenotypes at the seed-limited stage would cause a directional response in phenology due to confounding of those traits. Lastly, on-farm

trials were conducted in collaboration with the New York Corn and Soybean Growers Association to identify optimal planting rates for corn (*Zea mays* L.) and soybean (*Glycine max* L.) given the underlying spatial variability of the soil and topographical characteristics of the fields. A random forest regression-based approach was created to develop variable rate planting designs for maximizing yields.

BIOGRAPHICAL SKETCH

Margaret Krause grew up in Eden Prairie, Minnesota and attended the University of Minnesota in 2009. She graduated with a B.S. in Applied Plant Science in 2014 prior to transitioning to the Department of Plant Breeding and Genetics at Cornell University. Under the advisement of Drs. Mark Sorrells and Michael Gore, Margaret completed three relatively disparate projects at Cornell. During her first two years, she pursued a collaboration with the New York Corn and Soybean Grower's Association to develop strategies for implementing variable rate planting technologies for corn and soybean. In 2016, she became a visiting student at the International Maize and Wheat Improvement Center (CIMMYT) in México. For two years, she moved between the CIMMYT headquarters in El Batán and the Campo Experimental Norman E. Borlaug in Ciudad Obregón to assist the bread wheat breeding program and to conduct two projects on the integration of aerial high-throughput phenotyping and genomic selection to increase the rate of genetic gain. While at CIMMYT, Margaret was advised by Dr. José Crossa of the Biometrics and Statistics Unit. In September 2019, she will begin a postdoctoral fellowship funded by the National Science Foundation under the mentorship of Dr. Matthew Reynolds at CIMMYT and Dr. Jesse Poland at Kansas State University.

Dedicated to Judy, John and Mitch, Emily, Jimmy, Adrian, Jorge, Mitchell, and Chris.

ACKNOWLEDGEMENTS

Funding for this research was provided by several sources. My tuition and stipends were supported by the Cornell Fellowship, the New York Corn and Soybean Growers Association, the New York Soybean Checkoff, DuPont Pioneer, and the National Science Foundation Graduate Research Fellowship Program. The U.S. Borlaug Fellowship in Global Food Security provided supplemental support for my tenure as a visiting student at CIMMYT. Research support for my work with CIMMYT was provided in part by the Delivering Genetic Gain in Wheat project, funded by the U.K. Government's Department of International Development and the Bill & Melinda Gates Foundation. Partial support was also provided by the Agriculture and Food Research Initiative Competitive Grants Triticeae-CAP and Wheat-CAP from the United States Department of Agriculture National Institute of Food and Agriculture. Finally, additional support was provided by Feed the Future through the United States Agency for International Development.

I would like to acknowledge and thank my committee co-chairs Drs. Mark Sorrells and Michael Gore and my minor committee members Drs. Jessica Rutkoski and Susan Riha for their support and mentorship. I would also like to thank Savannah Crossman and Dr. José Crossa for their critical roles in facilitating my research with the New York Corn and Soybean Growers Association and CIMMYT, respectively. Finally, I would like to acknowledge Dr. Jesse Poland of Kansas State University for his valuable input into my research on aerial phenotyping.

Lastly I would like to express my deep gratitude for my friends and peers at Cornell University, the University of Minnesota, and CIMMYT, as well as those I met along the way in various corners of the globe. You have been instrumental to my academic and personal growth and have kept me afloat over the past five years.

TABLE OF CONTENTS

CHAPTER 2

HIGH-THROUGHPUT PHENOTYPING FOR INDIRECT SELECTION ON WHEAT GRAIN YIELD AT THE EARLY GENERATION SEED-LIMITED STAGE

CHAPTER 3

A RANDOM FOREST REGRESSION-BASED APPROACH FOR OPTIMIZING VARIABLE PLANTING RATES FOR CORN AND SOYBEAN

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

HYPERSPECTRAL REFLECTANCE-DERIVED RELATIONSHIP MATRICES FOR GENOMIC PREDICTION OF GRAIN YIELD IN WHEAT

Abstract

Hyperspectral reflectance phenotyping and genomic selection are two emerging technologies that have the potential to increase plant breeding efficiency by improving prediction accuracy for grain yield. Hyperspectral cameras quantify canopy reflectance across a wide range of wavelengths that are associated with numerous biophysical and biochemical processes in plants. Genomic selection models utilize genome-wide marker or pedigree information to predict the genetic values of breeding lines. In this study, we propose a multi-kernel GBLUP approach to genomic selection that uses genomic marker-, pedigree-, and hyperspectral reflectance-derived relationship matrices to model the genetic main effects and genotype × environment ($G \times E$) interactions across environments within a bread wheat (*Triticum aestivum* L.) breeding program. We utilized an airplane equipped with a hyperspectral camera to phenotype five differentially managed treatments of the yield trials conducted by the Bread Wheat Improvement Program of the International Maize and Wheat Improvement Center (CIMMYT) at Ciudad Obregón, México over four breeding cycles. We observed that single-kernel models using hyperspectral reflectance-derived relationship matrices performed similarly or superior to marker- and pedigree-based genomic selection models when predicting within and across environments. Multi-

kernel models combining marker/pedigree information with hyperspectral reflectance phenotypes had the highest prediction accuracies; however, improvements in accuracy over marker- and pedigree-based models were marginal when correcting for days to heading. Our results demonstrate the potential of using hyperspectral imaging to predict grain yield within a multi-environment context and also support further studies on the integration of hyperspectral reflectance phenotyping into breeding programs.

Introduction

The aim of plant breeding programs is to develop and deliver new, high-yielding crop varieties that are adapted to a range of environmental conditions. Two major challenges of breeding for multiple environments are: (1) being able to account for the presence of genotype-by-environment interaction ($G \times E$), and (2) the high costs of evaluating trials at multiple locations. Statistical, genomics, and phenomics tools that enable the accurate prediction and selection of candidate lines appropriate for each target environment may serve to increase the rate of genetic gain while reducing costs associated with large-scale multi-environment field trials.

In recent years, genomic selection (GS) and high-throughput phenotyping (HTP) have emerged as potential technologies for improving breeding efficiency (Furbank and Tester, 2011; Cabrera-Bosquet et al., 2012; Crossa et al., 2017). In GS, genome-wide marker effects are estimated for a "training set" of lines that has been phenotyped and genotyped (Meuwissen et al., 2001). Those estimates are then applied to selection candidates prior to phenotyping to predict trait values, which may reduce the time and cost of testing breeding lines. While GS was initially developed to predict within individual environments, recent studies have extended the genomic best linear

unbiased prediction (GBLUP) model to accommodate $G \times E$ interactions. Burgueño *et* al. (2012), Heslot et al. (2013), Jarquín et al. (2014), López-Cruz et al. (2015), and Pérez-Rodríguez et al. (2017) reported increases in prediction accuracies with extended models relative to single-environment analyses.

HTP is based on the remote and proximal sensing of a large number of crop plants to collect relevant phenotypes while reducing labor time and cost (White et al., 2012; Araus and Cairns, 2014; Pauli et al., 2016). When deployed across different developmental growth stages and in multiple environments, HTP can drastically increase the phenotypic information available to breeding programs, which may help to improve selection accuracy. Although HTP traits represent indirect estimations and may not be able to provide the level of precision of direct measurement, they may be of particular use at the early generation stage in breeding programs when seed is limited. In wheat breeding, large numbers of lines are sown in small, unreplicated plots for the purpose of visual selection and seed increase prior to grain yield testing in large replicated plots. Measurements of grain yield from small plots are not meaningful, nor are they feasible to collect when thousands of lines are being screened. Therefore, acquiring accurate predictions of grain yield at the early generation stage using HTP may serve to improve selection accuracy, though an initial proof of concept at the replicated yield trial stage is necessary to effectively develop prediction approaches.

A range of ground-based and aerial HTP platforms have been recently developed to improve the accuracy, efficiency, and scope of phenotypic data collection (Andrade-Sanchez et al., 2014; Crain et al., 2016; Haghighattalab et al.,

2016). A major advantage of aerial platforms is their ability to phenotype large areas of field trials in minimal time. This enhanced efficiency increases the spatial and temporal resolution of the phenotypic data and may be critical when assessing breeding trials at multiple locations. A number of recent studies have integrated traits collected with HTP into GS to increase prediction accuracy for grain yield in wheat (Rutkoski et al., 2014; Sun et al., 2017; Crain et al., 2018).

To date, many of the applications of aerial HTP within plant breeding programs have focused on measuring the spectral reflectance of the crop canopy. Plant cells, tissues, and pigments have wavelength-specific light absorption, reflectance, and transmittance patterns that may, for example, differentiate between healthy and stressed plants (Li et al., 2014). Vegetation indices (VIs) provide a convenient method to summarize spectral reflectance information into scores that may be predictive of economically important traits (Govaerts et al., 1999). VIs such as the normalized difference vegetation index (NDVI) have been shown to be predictive of grain yield in wheat (Aparicio et al., 1999; Labus et al., 2002). However, since VIs are calculated from only a few wavelengths, they cannot leverage the high density of information captured by hyperspectral cameras, which record reflectance at a large number of narrowband wavelengths in the visible and near infrared regions of the light spectrum (Viña et al., 2011). While hyperspectral data may have a greater capacity than VIs to detect phenotypic differences between individuals, the high dimensionality of the data may complicate the interpretation. For these data to become meaningful for plant breeding, methods that can derive useful information on traits relevant to breeding are needed.

To address the high dimensionality of hyperspectral data, Aguate et al. (2017) found that integrating information from all hyperspectral wavelengths using ordinary least squares, partial least squares, and Bayesian shrinkage resulted in higher prediction accuracy than what could be achieved using individual VIs in maize. Montesinos-López et al. (2017a) proposed a Bayesian functional regression analysis using hyperspectral wavelengths that likewise resulted in higher accuracies for predicting grain yield in wheat when compared to a range of VIs. Montesinos-López et al. (2017b) further extended this method to incorporate genomic and pedigree information, in addition to accommodating $G \times E$ by modeling hyperspectral band-by-environment ($B \times E$) interactions. Their study found that models that included the $B \times E$ term had higher prediction accuracies than those that did not, suggesting that hyperspectral reflectance may be a useful phenotype for modeling $G \times E$ interactions.

When collecting hyperspectral data within a multi-environment context, the number of predictors increases in proportion to the number of environments and phenotyping time-points observed, which may come at a computational cost depending on the type of prediction model used (Montesinos-López et al., 2017b). One possible approach that may minimize computation time would be to use the hyperspectral bands as a high dimensional predictor set, similar to the case of prediction with genomic markers in GBLUP, by creating a relationship matrix between individuals using the hyperspectral bands. This way, the number of bands could be very large without increasing the complexity of the GBLUP prediction model. Separate genomic marker/pedigree and hyperspectral reflectance kernels could be integrated to model the genetic main effects and $G \times E$ interactions, respectively.

Multi-environment field trials are often unbalanced, which can complicate their use in prediction across environments. When deploying HTP in large breeding programs, it can be difficult to ensure that all locations are phenotyped at the same stage and the same number of times throughout the season. Weather conditions, technical difficulties with the cameras or sensors, and scheduling with contracted pilots/airports may prohibit the use of aerial HTP on certain days, and lines grown at different sites may develop at faster or slower rates depending on weather and management conditions. As a result, sites may have different numbers of observed HTP time-points, as in Sun et al. (2017). It is well documented that canopy spectral reflectance varies according to crop phenology (Viña et al., 2003; Zhang et al., 2003). One strategy for comparing HTP time-points across varying sites may be to classify them according to the predominant developmental growth stage at the time of phenotyping.

To test these approaches, we deployed HTP on the CIMMYT Bread Wheat Improvement Program's multi-environment yield evaluations of advanced germplasm to phenotype a range of differentially managed treatments with a hyperspectral camera at multiple time-points throughout the growing season. The main objectives of this research were to: (1) propose a multi-kernel, multi-environment GBLUP model that involves modeling genetic main effects using genomic markers or pedigrees and modeling the $G \times E$ interactions using relationship matrices derived from hyperspectral reflectance data, (2) compare the prediction accuracies of models utilizing genomic marker/pedigree main effects kernels and hyperspectral $G \times E$ interaction kernels

separately and in combination, and (3) determine the optimal developmental growth stages for hyperspectral phenotyping based on grain yield prediction accuracy.

Materials and Methods

Experimental Data

The dataset included a total of 3,771 bread wheat lines evaluated at the Campo Experimental Norman E. Borlaug in Ciudad Obregón, México over the course of four breeding cycles: 2013-14, 2014-15, 2015-16, and 2016-17. In each cycle, lines were sown under five differentially managed treatments: Optimal Bed, Optimal Flat, Moderate Drought, Severe Drought, and Heat. Descriptions of the managed treatments are given in Table 1.1. The 20 managed treatment-breeding cycle combinations are herein referred to as site-years. Within each of the five managed treatments, 1,092 lines were arranged into 39 trials in a α-lattice design with three replicates and six incomplete blocks per replicate. Each replicate contained two repeated checks. 'Kachu #1' was sown as a check in the Optimal Bed, Optimal Flat, and Moderate Drought managed treatments. 'Baj #1' was used as a check in the Severe Drought and Heat managed treatments. 'Borlaug100 F2014' was sown as the other check in all managed treatments. All lines within a breeding cycle were evaluated in all five managed treatments, while no lines other than checks overlapped between cycles. Records for some lines were removed from the analysis due to unavailability of genotypic or agronomic data, resulting in the final dataset: 588 lines in 2013-14, 1,033 lines in 2014-15, 1,063 lines in 2015-16, and 1,087 lines in 2016-17, for a total of 56,565 plots phenotyped. Each breeding cycle contained full-sib families with an average of two full-sibs per family.

Table 1.1: Description of the field management conditions in each of the five managed treatments sown at the Campo Experimental Norman E. Borlaug in Ciudad Obregón, México

Managed Treatment	Planting Date	Plot Type	Plot Dimensions	Irrigation Methods
Optimal Bed	Late November/ Early December	Two beds with 3 rows per bed	2.8m × 0.8m	Five furrow irrigations
Optimal Flat	Late November/ Early December	Flat sown plot with 6 rows	4.0m × 1.3m	Five flood irrigations
Moderate Drought	Late November/ Early December	Two beds with 3 rows per bed	2.8m × 0.8m	Two furrow irrigations
Severe Drought	Late November/ Early December	Flat sown plot with 6 rows	4.0m × 1.3m	Three minimal irrigations through drip
Heat	Late February	Two beds with 3 rows per bed	2.8m × 0.8m	Five furrow irrigations

Grain yield (GY) in t ha^{-1} and lodging evaluated on an ordinal scale (0: no lodging; 5: completely lodged) were assessed in all three replicates. Heading date was recorded as the date of 50 percent spike emergence within the plot. Maturity date was recorded as the date of senescence of the peduncle for 50 percent of stems in the plot. Days to heading (DTHD) and days to maturity (DTMT) were measured as the number of days to reach heading and maturity, starting from the date of the first irrigation if sowing was on dry soil or from the sowing date if sowing was in pre-irrigated fields. DTHD and DTMT were assessed for the first replicate only due to the high heritability of those traits.

Hyperspectral Phenotyping

Hyperspectral reflectance data were collected with a hyperspectral camera (A-series, Micro-Hyperspec VNIR, Headwall Photonics, Fitchburg, MA, USA) as part of the Alava Remote Sensing Spectral Solution (ARS3, Alava Ingenieros, Madrid, Spain) mounted in a Piper PA-16 Clipper aircraft. The camera's sensor had a 12-bit radiometric resolution, covering the light spectrum in the 380–850nm region with a 7.5nm full width at half maximum and set with an integration time of 18ms. Spectral binning resulted in 62 wavelengths between 398-847nm (Rodrigues et al., 2018). The flights were scheduled around noon (GMT-7) and aligned to the solar azimuth angle at a height of 300m, resulting in 30cm resolution. The aerial imagery acquisition during the growing season was spaced approximately at seven- to ten-day intervals on mostly clear days.

Radiometric calibration of the sensor was done using coefficients derived from a calibrated uniform light source and an integrating sphere (CSTMUSS2000C

Uniform Source System, LabSphere, North Sutton, NH, USA). Dark frame correction was performed for each flight dataset. Atmospheric calibration was performed using irradiance measurements acquired at the beginning and end of each flight using a Jaz spectrometer with a CC-3 Cosine Corrector (Ocean Optics Inc, FL, USA) for the 2016-17 breeding cycle. For remaining three breeding cycles, irradiance was modeled using aerosol optical depth from sun-photometer measurements (Microtops II, Solar Light Company, Glenside, PA, USA) based on the SMARTS simulation model (Gueymard, 1995; Gueymard 2005).

Ortho-rectification and georeferencing of the imagery were performed using PARGE (ReSe Applications Schläpfer, Wil, Switzerland) based on data from the inertial navigation system (INS) attached to the camera (IG-500N model, SGB systems S. A. S., Carrières-sur-Seine, France). Hyperspectral reflectance data were extracted from the aerial images using the mean value of the pixels inside the central area of each observed plot, avoiding 0.5m from the plot border.

Each hyperspectral phenotyping time-point within each site-year was assigned a developmental growth stage classification according to the predominant growth stage of the lines within the site-year at the time of phenotyping (Table 1.2). The vegetative (VEG) stage was defined as the period between germination and 50 percent of plots at heading. The heading stage (HEAD) was defined as the period between 50 percent of plots at heading and 100 percent of plots at heading. The grain fill stage (GF) was defined as the period between 100 percent of plots at heading and 100 percent of plots at maturity.

Table 1.2: Classification of hyperspectral phenotyping time-points as vegetative (VEG), heading (HEAD), and grain fill (GF) phenological stages

Breeding	Phenotyping	Optimal	Optimal	Moderate	Severe	Heat
2013-14	10 Jan	-	VEG	VEG	-	-
	17 Jan	-	VEG	VEG	VEG	-
	30 Jan	VEG	VEG	VEG	VEG	-
	07 Feb	VEG	VEG	VEG	VEG	-
	14 Feb	HEAD	VEG	HEAD	HEAD	-
	19 Feb	HEAD	HEAD	HEAD	HEAD	-
	27 Feb	GF	HEAD	GF	GF	-
	11 Mar	GF	GF	GF	GF	-
	17 Mar	GF	GF	GF	-	-
	25 Apr	-	-	-	-	VEG
	02 May	-	-	-	-	HEAD
	07 May	-	-	-	-	HEAD
	21 May	-	-	-	-	GF
2014-15	10 Jan	VEG	VEG	VEG	VEG	-
	19 Jan	VEG	VEG	VEG	VEG	-
	04 Feb	VEG	VEG	VEG	VEG	-
	09 Feb	VEG	VEG	HEAD	HEAD	-
	25 Feb	HEAD	HEAD	GF	GF	-
	10 Mar	GF	GF	GF	GF	-
	15 Mar	GF	GF	GF	GF	-
	23 Mar	GF	GF	GF	-	-
	26 Mar	GF	GF	-	-	-
	07 Apr	-	GF	-	-	-
	14 Apr	-	-	-	-	VEG
	24 Apr	-	-	-	-	HEAD
	28 Apr	-	-	-	-	HEAD
	06 May	-	-	-	-	GF

Table 1.2 (Continued)

2015-16	26 Feb	HEAD	HEAD	GF	GF	-
	03 Mar	HEAD	HEAD	GF	GF	-
	09 Mar	GF	GF	GF	GF	-
	15 Mar	GF	GF	GF	GF	-
	22 Mar	GF	GF	GF	GF	-
	02 May	-	-	-	-	HEAD
	09 May	-	-	-	-	GF
	14 May	-	-	-	-	GF
2016-17	10 Jan	VEG	VEG	VEG	VEG	-
	17 Jan	VEG	VEG	VEG	VEG	-
	23 Jan	VEG	VEG	VEG	VEG	-
	02 Feb	VEG	VEG	VEG	HEAD	-
	10 Feb	VEG	VEG	HEAD	HEAD	-
	16 Feb	HEAD	VEG	HEAD	GF	-
	22 Feb	HEAD	HEAD	GF	GF	-
	15 Mar	GF	GF	-	GF	-
	29 Mar	GF	-	-	-	-
	05 Apr	GF	-	-	-	-
	11 Apr	GF	-	-	-	-
	02 May	-	-	-	-	HEAD
	22 May	-	-	-	-	GF
	31 May	-	-	-	-	GF

Genotypic Data

All lines were genotyped using genotyping-by-sequencing (Elshire et al., 2011) according to the pipeline described in Poland et al. (2012). From the initial set of 34,900 single nucleotide polymorphisms (SNPs), 8,519 SNPs remained after excluding all markers with more than 70 percent of missing data or minor allele frequency less than 0.05. For each marker, missing data were imputed using the sample mean of observed values (Poland et al., 2012).

Basic Statistical Models

Within each site-year, best linear unbiased estimates (BLUEs) were calculated for the agronomic traits and for each hyperspectral band at each phenotyping time-point using the following model:

$$y_{ijkl} = \mu + g_i + t_j + r_{k(j)} + b_{l(jk)} + \varepsilon_{ijkl} \quad (1)$$

where y_{ijkl} is the trait value for genotype i within trial j, replicate k, and block l; μ is the overall mean; g_i is the fixed effect for genotype i; t_j is the random effect for trial j, which are assumed to be independently and identically distributed according to a normal distribution with mean zero and variance σ_t^2, that is, $t_j \sim$iid $N(0, \sigma_t^2)$; $r_{k(j)} \sim$iid $N(0, \sigma_r^2)$ is the random effect for replicate k within trial j; $b_{l(jk)} \sim$iid $N(0, \sigma_b^2)$ is the random effect for block l within replicate k and trial j; and $\varepsilon_{ijkl} \sim iid\ N(0, \sigma_\varepsilon^2)$ is the residual effect. For DTHD and DTMT, which were recorded for only one replicate, the replicate and block effects were excluded from the model. For validation of prediction models, best linear unbiased predictions (BLUPs) for GY within each site-year were calculated by treating the effect for genotype g_i in model (1) as random with $g_i \sim iid\ N(0, \sigma_g^2)$ and by including a covariate for lodging as a fixed effect, fit only

when lodging was observed within a site-year. To correct for the influence of phenology, BLUPs for GY were calculated again using DTHD included as a fixed effect in model (1).

BLUEs were also calculated for each hyperspectral band on a developmental growth stage basis according to the growth stage classifications detailed in Table 1.2. BLUEs for each hyperspectral band for the VEG, HEAD, and GF stages were estimated by fitting the following model:

$$y_{ijklm} = \mu + g_i + d_j + t_{k(j)} + r_{l(jk)} + b_{m(jkl)} + \varepsilon_{ijklm} \quad (2)$$

where y_{ijkl} is the hyperspectral reflectance trait value for genotype i on time-point j and within trial k, replicate l, and block m; μ is the overall mean; g_i is the fixed genetic effect for genotype i; $d_j \sim iid\ N(0, \sigma_d^2)$ is a random effect for time-point j; $t_{k(j)} \sim iid\ N(0, \sigma_t^2)$ is random effect for trial k nested within time-point j; $r_{l(jk)} \sim iid\ N(0, \sigma_r^2)$ is the random effect for replicate l nested within time-point j and trial k; $b_{m(jkl)} \sim$iid $N(0, \sigma_b^2)$ is the random effect for block m nested within time-point j, trial k, and rep l; and $\varepsilon_{ijklm} \sim$iid $N(0, \sigma_\varepsilon^2)$ is the residual effect. BLUEs were also calculated for each hyperspectral band across all available phenotyping time-points (ALL) using model (2).

Broad-sense heritability within site-years was calculated for GY (Bernardo 2010) and each hyperspectral band as:

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_\varepsilon^2}{nreps}} \quad (3)$$

where σ_g^2 is the genetic variance, σ_ε^2 is the error variance, and *nreps* is the number of replicates (*nreps* = 3). σ_g^2 and σ_ε^2 were derived by fitting model (1) with the effect for

genotype g_i treated as random with $g_i \sim iid\ N(0, \sigma_g^2)$. For each breeding cycle, Pearson's correlations for GY between managed treatments were calculated using the BLUEs derived from model (1). Pearson's correlations were also calculated between GY BLUEs and the BLUEs for each hyperspectral band calculated in model (2).

Relationship Matrices

Genetic relationships between individuals were modeled using genomic markers, pedigrees, and hyperspectral reflectance phenotypes. The genomic relationship matrix (***G***) was calculated according to Endelman and Jannink (2012). The additive relationship matrix (***A***) was derived from pedigrees and calculated as twice the coefficient of parentage.

Hyperspectral reflectance-based relationship matrices (***H***) were calculated within each site-year using the hyperspectral BLUEs calculated for 1) the individual time-points from model (1), 2) the developmental growth stages from model (2), and 3) all time-points from model (2). The matrices were calculated as $\boldsymbol{H} = \frac{\boldsymbol{SS'}}{nbands}$, where ***S*** is a matrix of the centered and standardized BLUEs of the hyperspectral bands and *nbands* is the total number of hyperspectral bands observed (*nbands* = 62).

Prediction Models

Genetic Main Effects: To compare the utility of hyperspectral reflectance-based models to standard marker- and pedigree-based methods for genomic prediction, the following single-kernel genetic main effects model was fitted using marker- and pedigree-derived relationship matrices:

$$y_{ij} = \mu + E_i + g_j + \varepsilon_{ij} \quad (4)$$

where y_{ij} is the BLUE of GY for genotype j in site-year i, μ is the overall mean, E_i is the fixed effect for site-year ($i=1,\ldots,I$), g_j is the random effect for genotype j ($j=1,\ldots,J$), and ε_{ij} is the residual effect. We assume that the joint distribution of genotype effects is distributed according to a multivariate normal distribution with mean $\mathbf{0}$ and variance-covariance matrix $\sigma_g^2\boldsymbol{K}$, that is, $\boldsymbol{g} = (g_1, \ldots, g_J)^T \sim MVN(\mathbf{0}, \sigma_g^2\boldsymbol{K})$, where σ_g^2 denotes the genomic variance and $\boldsymbol{K}$ represent the genomic relationship matrix between individuals $\boldsymbol{G}$ ($\boldsymbol{K}$=$\boldsymbol{G}$) or the additive pedigree relationship matrix $\boldsymbol{A}$ ($\boldsymbol{K}$=$\boldsymbol{A}$).

Hyperspectral Reflectance Main Effects: To mimic a situation in which genomic marker and pedigree information are not available, the following single-kernel model was fitted using the hyperspectral reflectance-derived relationship matrices only:

$$y_{ij} = \mu + E_i + Eh_{ij} + \varepsilon_{ij} \quad (5)$$

where y_{ij}, μ, and E_i are as defined in model (4). Eh_{ij} is the random effect of the hyperspectral bands for genotype j in site-year i with the joint distribution of the hyperspectral bands as $\boldsymbol{hE} = (hE_{11}, \ldots, hE_{IJ})^T \sim MVN(\mathbf{0}, \sigma_h^2\boldsymbol{H})$, where σ_h^2 is the hyperspectral band variance and $\boldsymbol{H}$ is the hyperspectral reflectance-derived relationship matrix. Unlike matrices $\boldsymbol{G}$ and $\boldsymbol{A}$, which have only one row and column for each unique genotype evaluated, the matrix $\boldsymbol{H}$ has a row and column for each unique site-year-genotype combination where $I \times J$ results in the total number of site-year-genotype combinations.

Genetic Main Effects + Genetic G × E: As a basis for comparison to assess the advantage of $G \times E$ models that integrate both marker or pedigree relationship matrices with hyperspectral reflectance-derived relationship matrices, model (4) was extended

to accommodate $G \times E$ interactions using marker or pedigree information. The following multi-kernel model was fit:

$$y_{ij} = \mu + E_i + g_j + gE_{ij} + \varepsilon_{ij} \quad (6)$$

where y_{ij}, μ, E_i, and g_j are as defined in model (4). The gE_{ij} term is assumed to have multivariate normal distribution $\boldsymbol{gE} = (gE_{11}, \ldots, gE_{IJ})^T \sim MVN(\mathbf{0}, (\mathbf{Z_g K Z_g}^{\mathrm{T}}) \circ (\mathbf{Z_E Z_E}^{\mathrm{T}}) \sigma^2_{gE})$ where $\mathbf{Z_g}$ and $\mathbf{Z_E}$ are incidence matrices for genotypes and site-years, $\boldsymbol{K}$ represents the genomic relationship matrix ($\boldsymbol{K}$=$\boldsymbol{G}$) or the additive pedigree relationship matrix $\boldsymbol{A}$ ($\boldsymbol{K}$=$\boldsymbol{A}$), and σ^2_{gE} is the variance component for gE_{ij} (Jarquín et al., 2014).

Genetic Main Effects + Hyperspectral Reflectance G × E: Finally, a multi-kernel model using marker or pedigree information to estimate the genetic main effects and hyperspectral reflectance phenotypes to model the $G \times E$ interactions was fitted:

$$y_{ij} = \mu + E_i + g_j + hE_{ij} + \varepsilon_{ij} \quad (7)$$

Here, y_{ij}, μ, E_i, and g_j are defined as above in model (4). The term hE_{ij} is assumed to have multivariate normal distribution $\boldsymbol{hE} = (hE_{11}, \ldots, hE_{IJ})^T \sim MVN(\mathbf{0}, (\mathbf{Z_g H Z_g}^{\mathrm{T}}) \circ (\mathbf{Z_E Z_E}^{\mathrm{T}}) \sigma^2_{hE})$ where $\mathbf{Z_g}$ and $\mathbf{Z_E}$ are incidence matrices for genotypes and site-years and σ^2_{hE} is the variance component for hE_{ij}. The $(\mathbf{Z_g H Z_g}^{\mathrm{T}}) \circ (\mathbf{Z_E Z_E}^{\mathrm{T}})$ term is obtained with the block diagonal matrix $BDiag(\boldsymbol{H}_{11},\ldots,\boldsymbol{H}_{II})$ where $\boldsymbol{H}_{ii}$ represent the hyperspectral relationship matrices for genotypes in site-year $i=1,\ldots,I$.

Software

Processing of the hyperspectral images was performed using the ARS3 hyproQ software (Álava Ingenieros, Madrid, Spain). Plot polygons for tabular data extraction

were generated using ArcGIS (ESRI, Redlands, California, USA). Images were aligned manually in ArcGIS if they did not overlay the plot polygons due to INS inaccuracy.

All models were fit using the R statistical programming language (R Core Team, 2018). Basic models (1-3) were fit with the package "ASReml-R" (Gilmour et al., 2014) for R, while the prediction models (4-7) were fit using the "BGLR" package (de los Campos and Pérez-Rodríguez, 2014) for R. The marker-based genetic relationship matrices were calculated using the A.mat() function within the "rrBLUP" package (Endelman 2011) for R. The coefficients of parentage for the pedigree relationship matrices were estimated using the "Browse" application within the International Crop Information System software package (McLaren et al., 2000).

Assessing Model Prediction Accuracy

The above models were fit to assess model prediction accuracy for three prediction strategies representing different testing and evaluation problems relevant to plant breeding programs: 1) **within site-year**, 2) **within breeding cycle/across managed treatments**, and 3) **across breeding cycles/within managed treatment**.

A train-test (TRN-TST) validation scheme (Daetwyler et al., 2013) of 20 random partitions was used to assess model prediction accuracy for all prediction strategies. For within site-year prediction, a random 80 percent of records within a given site-year were assigned to the TRN set, and the remaining 20 percent were used as the TST set for prediction. This reflects a scenario in which prediction of the trait values for a set of unobserved lines is performed within a managed treatment and breeding cycle of interest. Models (4) and (5) were fitted without the E_i term for site-

year and models (6) and (7) were not tested for the within site-year prediction scheme because multiple site-years were not considered.

For prediction within breeding cycle/across managed treatments, predictions were carried out across the five managed treatments within a single breeding cycle. The TRN set consisted of all records from four out of the five managed treatments within a breeding cycle plus 20 percent of records from the fifth managed treatment. The remaining 80 percent of records from the fifth managed treatment were assigned to the TST set for prediction. In this situation, lines have been previously characterized in all managed treatments except one, where only 20 percent of records were available.

For prediction across breeding cycles/within managed treatment, predictions were performed across the four breeding cycles but within a single managed treatment. The TRN set contained all records from three out of the four breeding cycles for a given managed treatment plus 20 percent of records from the fourth breeding cycle. The remaining 80 percent of records from the fourth breeding cycle were assigned to the TST set for prediction. This scenario predicts performance of unobserved lines that are 1 to 3 breeding cycles removed from lines in the training set.

Table S1.1 illustrates examples of TRN-TST partitioning for the three prediction schemes. For all prediction strategies, model prediction accuracy was assessed as the Pearson's correlation between predictions for GY and GY BLUPs with and without correction for DTHD calculated in model (1). Reported values are the mean and standard deviation of the 20 random TRN-TST partitions implemented. The same partitioning "splits" were used to assess all models to ensure fair comparisons.

Data Availability

All phenotypic and genotypic data required to confirm the results presented in this study are available on CIMMYT Dataverse link: hdl:11529/10548109. The "GID" column denotes the unique identifiers for the genotypes. Supplementary file sets, tables and figures are available at FigShare.

Results

Descriptive Statistics

To evaluate the potential of integrating aerial hyperspectral reflectance phenotypes into GS to improve prediction accuracy for GY in wheat, we deployed an airplane equipped with a hyperspectral camera to phenotype wheat breeding trials in 20 site-years at multiple time-points throughout the growing season. Across the site-years, the Optimal Bed and Optimal Flat managed treatments had higher GY than the stressed Moderate Drought, Severe Drought, and Heat managed treatments (Figure 1.1). The standard deviations for GY remained relatively stable across site-years, ranging from 0.30 to 0.68 t ha^{-1} (Table S1.2). Broad-sense heritability for GY was high across site-years, ranging from 0.58 to 0.94 (Table S1.2). Overall, correlations for GY between managed treatments were moderately positive (Figure S1.1). The managed treatments that received similar levels of irrigation (e.g., Optimal Bed with Optimal Flat; Moderate Drought with Severe Drought) showed significant correlations (p-value ≤ 0.05) in all breeding cycles.

Between 3 and 11 hyperspectral phenotyping time-points were collected within each site-year (Table 1.2). In most site-years, at least one hyperspectral phenotyping time-point was collected during each of the three developmental growth stages. While

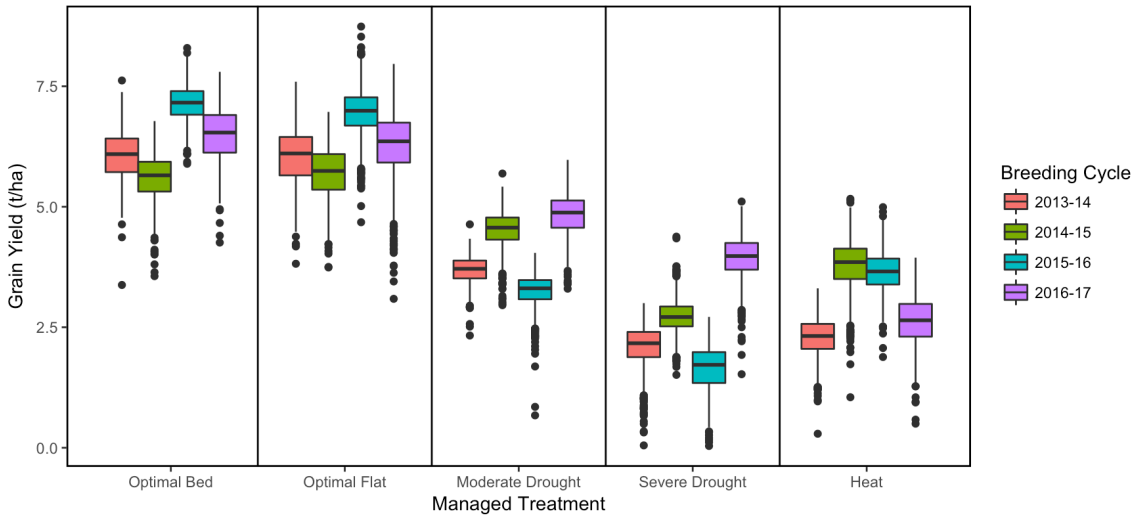


Figure 1.1: Boxplot of BLUEs for wheat grain yield (t ha^{-1}) in each of the 20 observed site-years

the managed treatments within a breeding cycle were often phenotyped on the same date, the growth stage at the time of phenotyping frequently differed among the managed treatments (Figure 1.2). During the 2015-16 cycle, technological issues with the camera prevented early season phenotyping. As a result, no hyperspectral data were collected during the VEG stage. The Optimal Bed, Optimal Flat, and Heat managed treatments had hyperspectral phenotypes for the HEAD and GF stages only, while the Moderate Drought and Severe Drought managed treatments had hyperspectral phenotypes for the GF stage only due to early maturation of the crop under water stress.

Broad-sense heritability estimates for the hyperspectral bands were generally between 0.5 and 0.8 for most phenotyping time-points (Figure 1.3). Heritabilities were relatively homogeneous across individual time-points and developmental growth stages. In all site-years, the lowest heritabilities were observed for the hyperspectral bands between 398 and 425 nm. Heritabilities in the 2013-14 Optimal Flat site-year were notably lower than in other site-years, falling between 0.00 and 0.25, though the GF time-points were more heritable with values around 0.40. Low heritabilities (<0.20) were also observed for the 9 March time-point in the 2015-16 Moderate Drought site-year.

Correlations between the individual hyperspectral bands and GY ranged from -0.68 to 0.64 across the 20 site-years and were most frequently the strongest above 500 nm (Figure 1.4). While there were some observable similarities in correlation patterns among time-points taken during the same growth stage within the same site-year, patterns were not uniform across site-years, breeding cycles, or managed treatments.

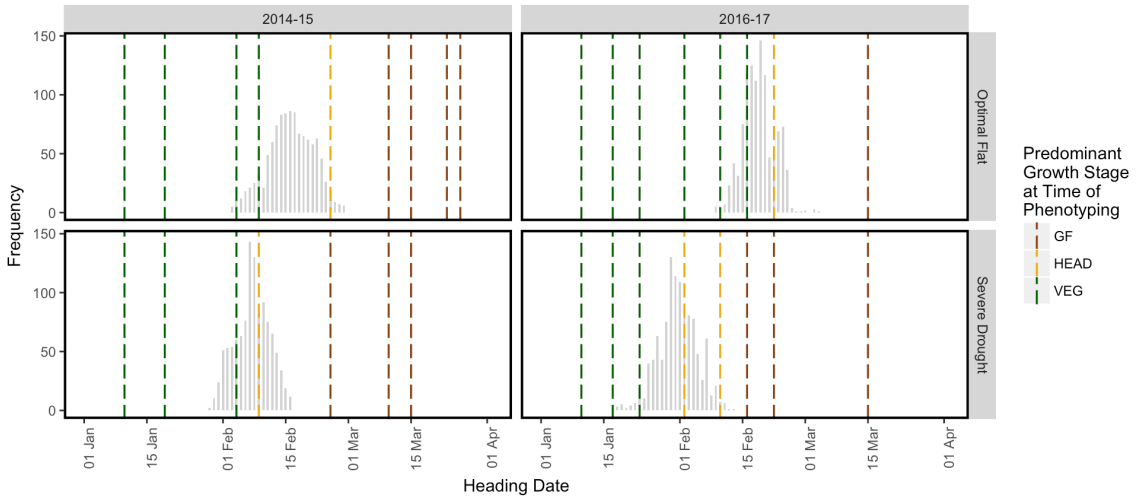


Figure 1.2: A graphical representation of the unbalanced nature of the hyperspectral reflectance phenotypic data

Four site-years are represented: 2014-15 Optimal Flat, 2014-15 Severe Drought, 2016-17 Optimal Flat, and 2016-17 Severe Drought. The histograms represent heading dates. Each dashed line corresponds to a hyperspectral phenotyping date colored according to the predominant growth stage of the lines at the time of phenotyping.

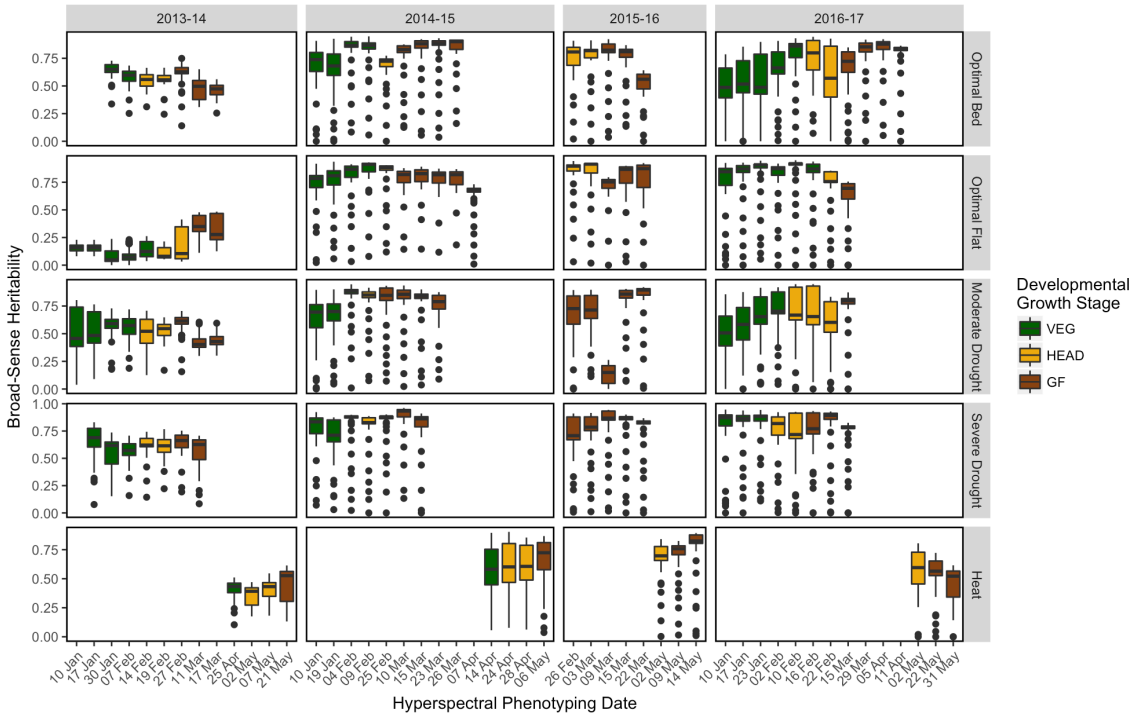


Figure 1.3: Broad-sense heritabilities of the hyperspectral wavelengths for each phenotyping time-point within each site-year

Each boxplot represents the distribution of broad-sense heritability values for the 62 hyperspectral wavelengths observed. The colors correspond to the developmental growth stage of the site-year at the time of hyperspectral phenotyping.

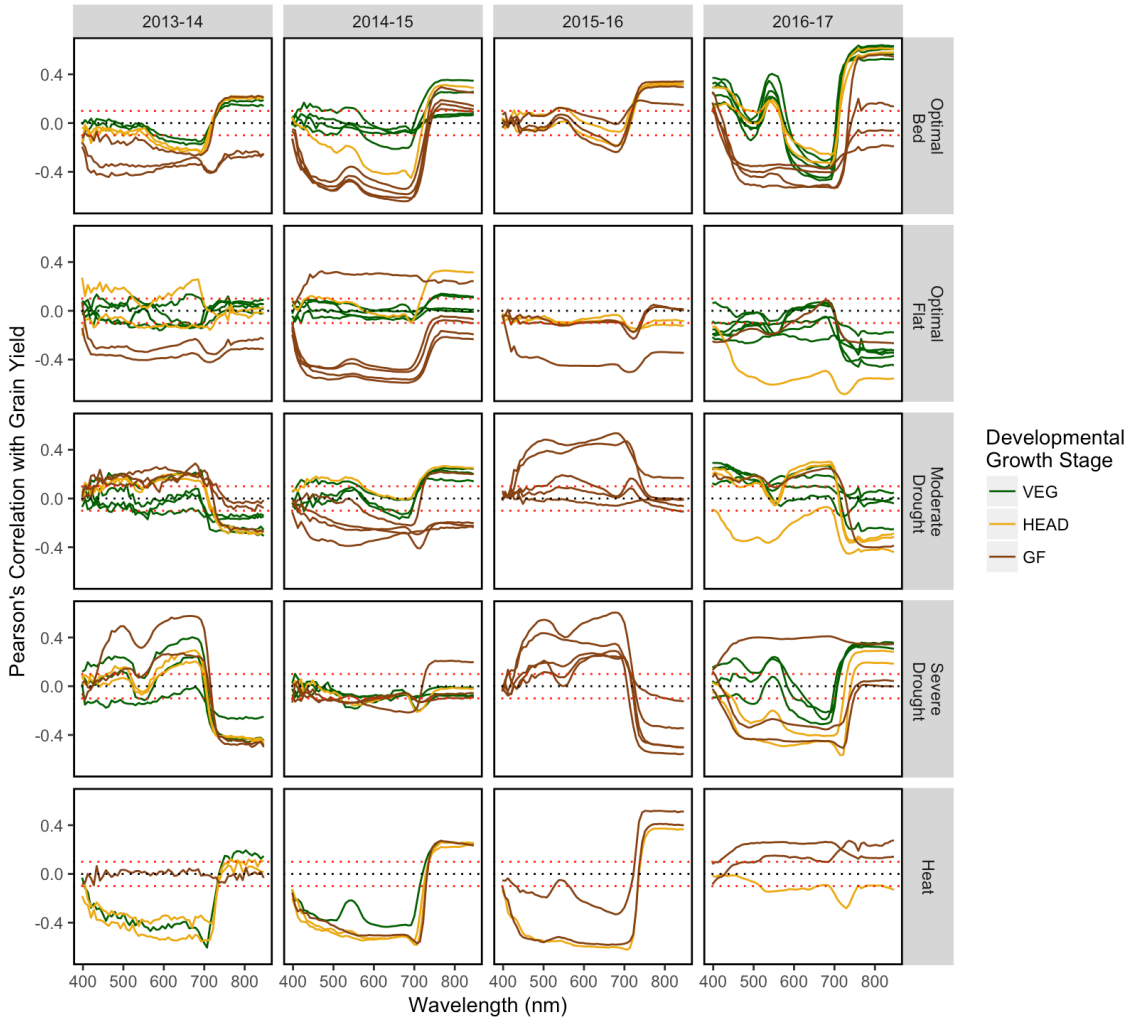


Figure 1.4: Empirical correlations between grain yield and hyperspectral reflectance BLUEs within each site-year

Each solid line represents a phenotyping time-point and the Pearson's correlation between grain yield and the 62 hyperspectral wavelengths observed. Lines are colored according to the predominant developmental growth stage of the site-year at the time of hyperspectral phenotyping. Correlation values ≥ |0.10| are significant at a level of 0.001, as denoted by the dotted red lines.

In some site-years, there were clear differences in correlation patterns between developmental growth stages. For example, in the 2014-15 Optimal Bed site-year, correlations between GY and hyperspectral reflectance in the 398 to 700 nm range were around 0 during the VEG stage but became progressively more negative during the HEAD and GF stages, reaching around -0.4. For the same managed treatment during the 2016-17 breeding cycle, distinct differences in correlation patterns between the three developmental growth stages were also observed; however, the patterns do not reflect those observed in 2014-15. Despite irregularities in correlation patterns, the correlations between hyperspectral bands from the GF stages and GY were, on average, stronger by 0.10 than for hyperspectral phenotypes taken during the VEG and HEAD stages.

Model Prediction Accuracy

Within Site-Year: Five types of relationship matrices were used within single-kernel models (4) and (5) for within site-year prediction: genomic (***G***), pedigree (***A***), individual hyperspectral phenotyping time-points (e.g. ***H.10Jan***, ***H.23Mar***, etc.), hyperspectral phenotype BLUEs calculated from multiple time-points for each developmental growth stage (***H.VEG***, ***H.HEAD***, ***H.GF***), and hyperspectral phenotype BLUEs calculated using all available time-points (***H.ALL***). Model (4) was used to assess accuracies using the ***G*** and ***A*** relationship matrices for prediction. Model (5) was fit to assess accuracies using the hyperspectral reflectance-derived relationship matrices.

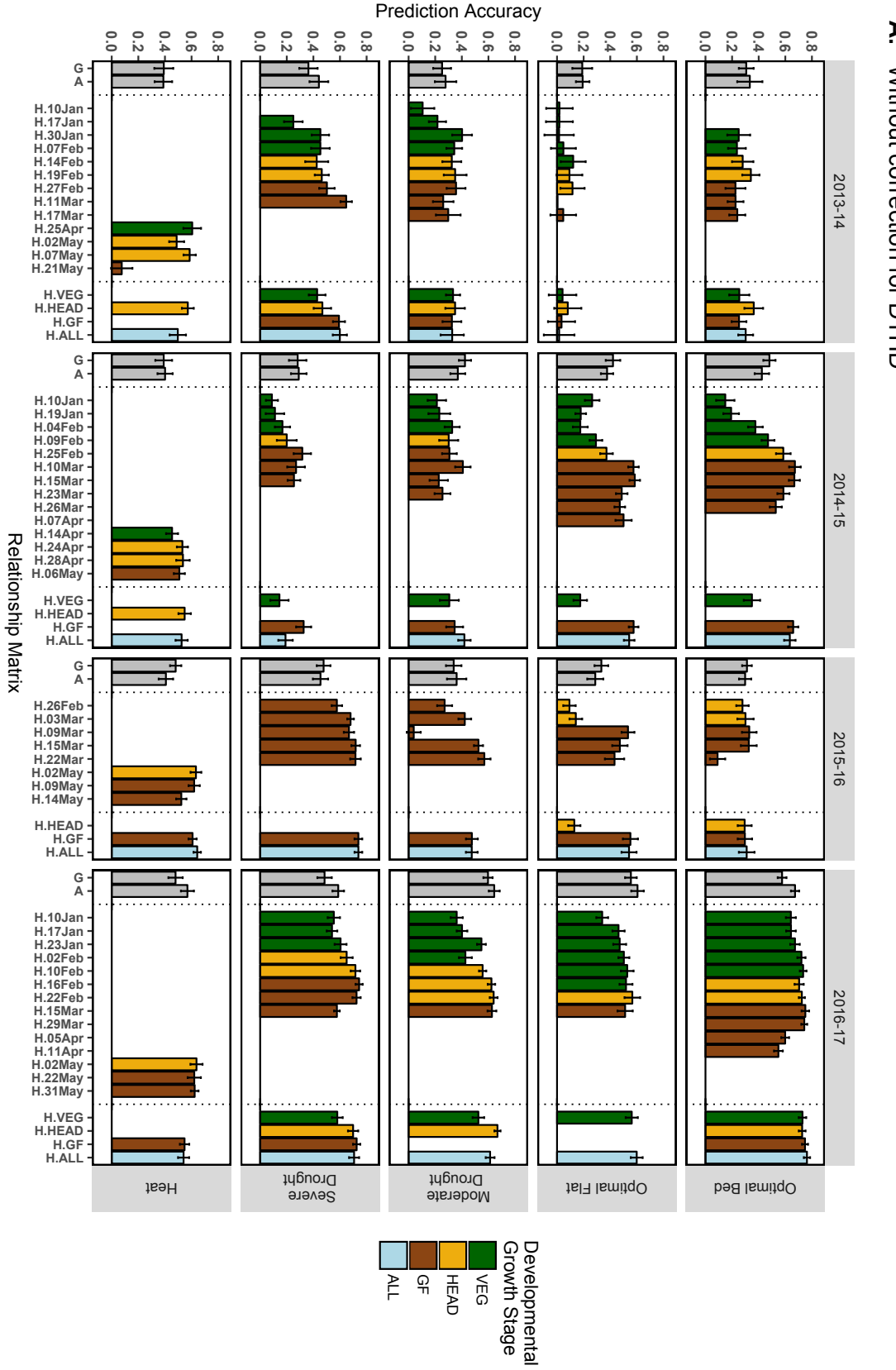
Prediction accuracies of models using the individual hyperspectral time-point-derived relationship matrices ranged from 0.00 to 0.75 with a mean of 0.42 (Figure

1.5A). These results are similar to the prediction accuracies observed for models using the ***G*** and ***A*** relationship matrices, which recorded means of 0.41 and 0.42 and ranges of 0.19 to 0.60 and 0.19 to 0.67, respectively. In 18 and 17 of the 20 site-years, at least one hyperspectral phenotyping time-point showed accuracies that were equivalent or superior to ***G*** and ***A***, respectively. In considering the most highly predictive hyperspectral phenotyping time-point in each site-year, 11 were recorded during the GF stage, followed by 6 and 3 in the HEAD and VEG stages, respectively, although the site-years had different numbers of hyperspectral phenotyping time-points from each growth stage and some site-years had no observations recorded during the VEG and HEAD stages.

The level of prediction accuracy for the individual hyperspectral time-points was highly correlated with the strength of the relationship between hyperspectral reflectance and GY (Figure 1.6). For a given time-point within a site-year, the correlation between hyperspectral reflectance and GY was taken for each of the 62 hyperspectral wavelengths. The average of the absolute values of those correlations, or their relative magnitudes, was then compared to the level of prediction accuracy for that time-point. For the 71 time-point/site-year combinations, the average magnitude of the correlation between hyperspectral reflectance and GY explained 51 percent of the variation in prediction accuracy, though the two characters exhibited a non-linear relationship (Figure 1.6A). Correlations beyond 0.25 between hyperspectral reflectance and GY provided little marginal increase in prediction accuracy. When GY was corrected for DTHD, this relationship was linear, and the correlation between

Figure 1.5: Within site-year prediction accuracies, with and without correction for DTHD

A. Without correction for DTHD



B. With correction for DTHD

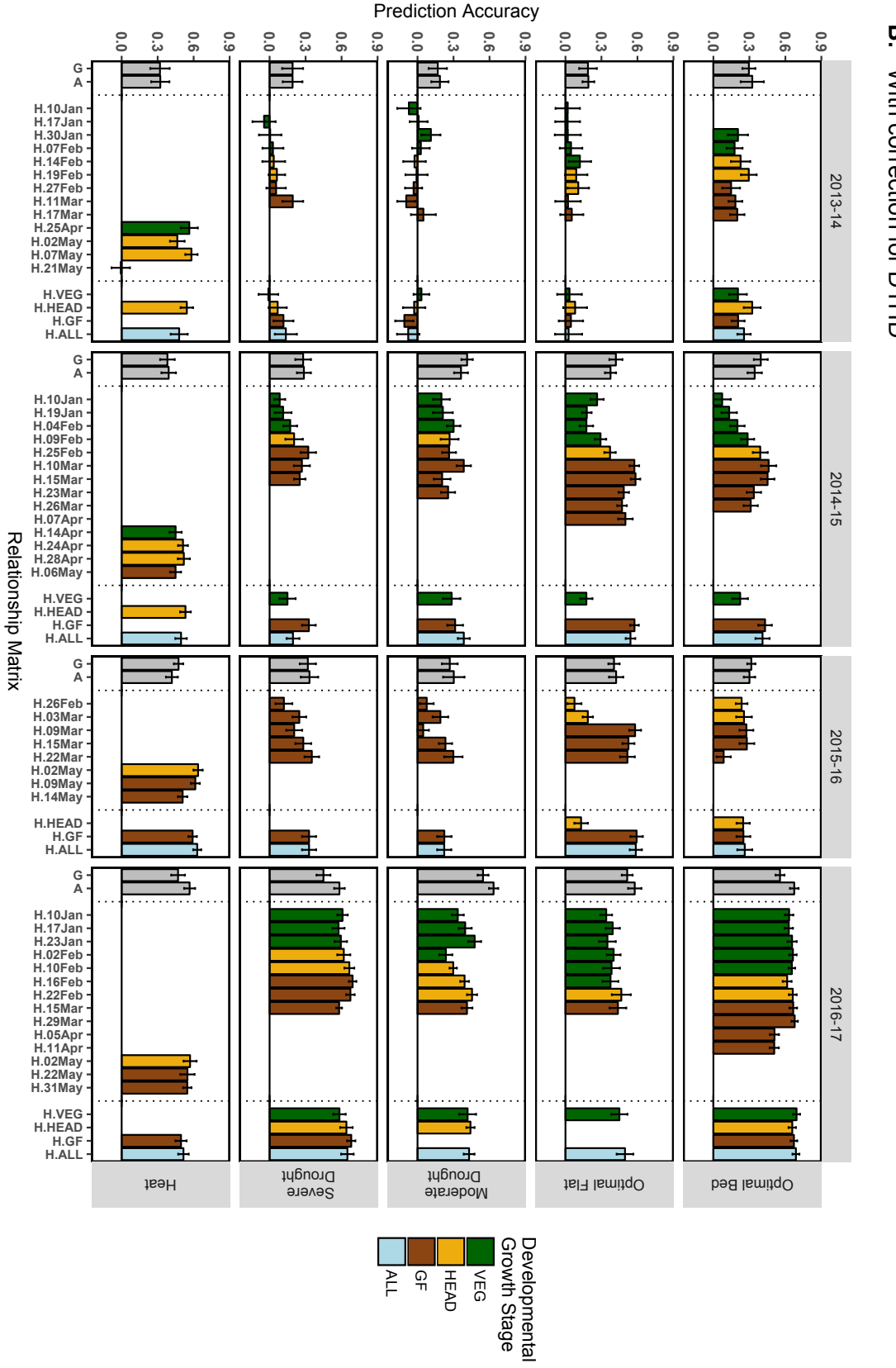
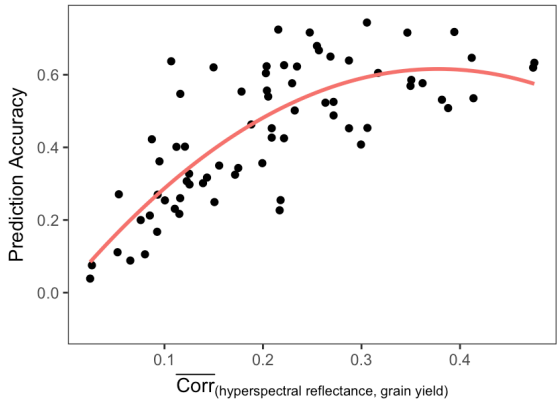


Figure 1.5 (Continued)

Accuracy is expressed as the average Pearson's correlation between predictions and observed BLUPs for GY across 20 random TRN-TST partitions. Results shown according to the type of relationship matrix tested: Genomic (**G**), pedigree (**A**), individual hyperspectral time-points (e.g. **H.10Jan**, **H.23Mar**, etc.), hyperspectral BLUES for each developmental growth stage (**H.VEG**, **H.HEAD**, **H.GF**), and hyperspectral BLUEs across all time-points (**H.ALL**). The color corresponds to the predominant developmental growth stage of the site-year at the time of phenotyping. Error bars are the standard deviation of prediction accuracy for the 20 random partitions.

A. WITHOUT DTHD CORRECTION:



B. WITH DTHD CORRECTION:

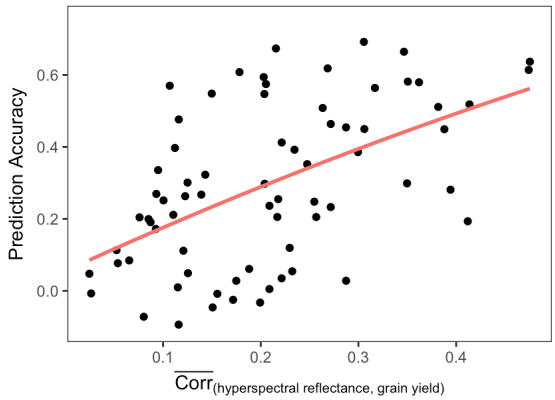


Figure 1.6: The relationship between individual hyperspectral time-point prediction accuracy and the average magnitude of the correlations between hyperspectral bands and GY

The absolute values of the correlations between hyperspectral bands and GY were calculated and then averaged across the 62 bands for each time-point within each site-year. Plotted on the x-axis, this represents the average strength of the relationship between hyperspectral reflectance and GY for each time-point within each site-year. The y-axis shows the prediction accuracy for each individual hyperspectral time-point in within site-year prediction.

reflectance and GY explained 26 percent of the variation in prediction accuracy (Figure 1.6B).

In considering prediction models using relationship matrices derived from the hyperspectral phenotype BLUEs for each growth stage, relationship matrix ***H.GF*** had the highest prediction accuracies in 11 of the 20 site-years, followed by ***H.HEAD*** in 8 site-years and ***H.VEG*** in 1 site-year. Combining multiple time-points in this manner did not necessarily increase accuracy. The accuracies of models using relationship matrices ***H.VEG***, ***H.HEAD***, and ***H.GF*** were comparable to the mean of the accuracies of models using individual time-point-derived relationship matrices. Likewise, the accuracies of models using relationship matrix ***H.ALL***, which combined all available time-points, were on average slightly higher than the mean of the accuracies of the individual time-point models but lower than models using the most predictive time-point.

While correcting for DTHD had negligible impacts on prediction accuracy for models using the ***G*** and ***A*** relationship matrices, the accuracies of models using the hyperspectral reflectance-derived relationship matrices decreased by 0.10 on average (Figure 1.5B). The greatest reductions were observed in the Severe Drought and Moderate Drought treatments, which showed average decreases in accuracy of 0.20 and 0.18, respectively. After the DTHD correction, at least one individual time-point-derived relationship matrix recorded prediction accuracies that were equivalent or superior to accuracies when using relationship matrices ***G*** and ***A*** in 13 out of the 20 site-years.

Within Breeding Cycle/Across Managed Treatments: For prediction across managed treatments within a breeding cycle, the genomic-, pedigree-, and hyperspectral reflectance-derived relationship matrices were tested individually in single-kernel models and in combination in multi-kernel models (Table 1.3). Prediction accuracies are shown in Figure 7. Single-kernel models included the following: genetic main effects (***G*** and ***A***) from model (4) and hyperspectral reflectance main effects (***H.VEG***, ***H.HEAD***, ***H.GF***, and ***H.ALL***) from model (5). Multi-kernel models were built by combining a main effects kernel with a $G \times E$ interaction kernel in models (6) and (7). In model (6), the $G \times E$ interaction kernel was also fit using relationship matrices ***G*** or ***A***. These models are herein referred to as $\boldsymbol{G} + \boldsymbol{G}_{\boldsymbol{GxE}}$ and $\boldsymbol{A} + \boldsymbol{A}_{\boldsymbol{GxE}}$. In model (7), the hyperspectral reflectance-derived relationship matrices were used to estimate the $G \times E$ interaction kernel. These models are referred to as $\boldsymbol{G} + \boldsymbol{H.VEG}_{\boldsymbol{GxE}}$, $\boldsymbol{A} + \boldsymbol{H.HEAD}_{\boldsymbol{GxE}}$, etc. according to the respective relationship matrices used to estimate the genetic main effects and $G \times E$ interactions. The ***H.VEG***, $\boldsymbol{G} + \boldsymbol{H.VEG}_{\boldsymbol{GxE}}$, and $\boldsymbol{A} + \boldsymbol{H.VEG}_{\boldsymbol{GxE}}$ models were assessed for 14 out of the 20 site-years due to the unavailability of hyperspectral data at the vegetative stage in the remaining 6 site-years. Likewise, the ***H.HEAD***, $\boldsymbol{G} + \boldsymbol{H.HEAD}_{\boldsymbol{GxE}}$, and $\boldsymbol{A} + \boldsymbol{H.HEAD}_{\boldsymbol{GxE}}$ models were assessed for 18 out of the 20 site-years. The remaining models using relationship matrices ***H.GF*** and ***H.ALL*** were tested for all site-years. When considering the single-kernel hyperspectral main effect models, 2013-14 Optimal Flat had prediction accuracies close to zero. The 2013-14 Optimal Flat site-year had high levels of lodging, which may have affected the reflectance signatures of the crop canopy. The hyperspectral reflectance phenotypes for 2013-14 Optimal Flat

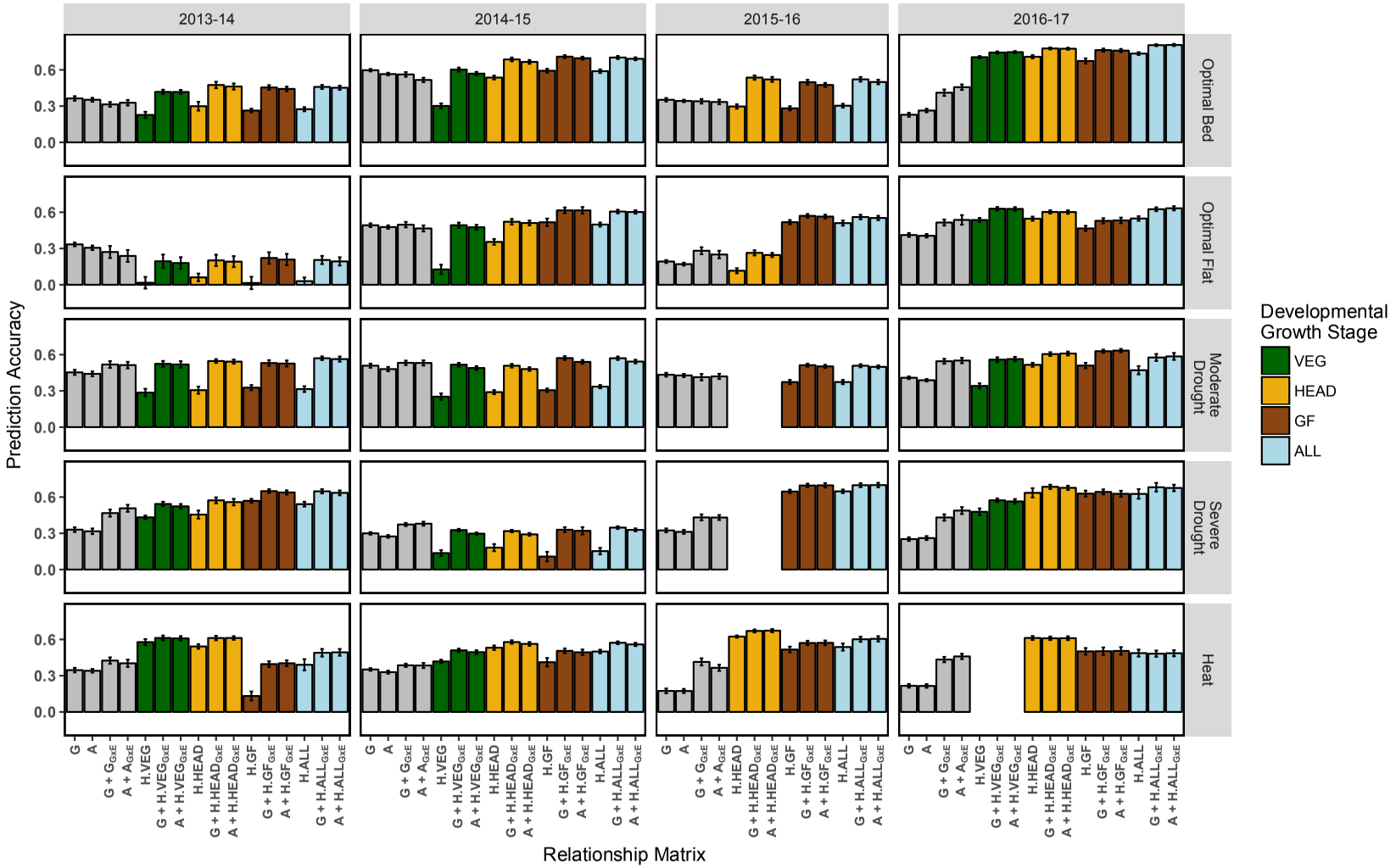
Table 1.3: Description of the single- and multi-kernel models tested for prediction within breeding cycle/across managed treatments and across breeding cycles/within managed treatment

Model Abbreviation	Relationship Matrix Used to Model the Genetic Main Effect			Relationship Matrix Used to Model the G×E Interaction Effect		
	G	***A***	***H.VEG***	***G***	***A***	***H.VEG***
G	X					
A		X				
H.VEG			X			
$G + G_{GxE}$	X			X		
$A + A_{GxE}$		X			X	
$G + H.VEG_{GxE}$	X					X
$A + H.VEG_{GxE}$		X				X

G is the genomic relationship matrix, ***A*** is the additive pedigree relationship matrix, and ***H.VEG*** is the hyperspectral relationship matrix derived from BLUEs calculated across hyperspectral phenotyping time-points collected at the vegetative stage. Single- and multi-kernel models using relationship matrices ***H.HEAD***, ***H.GF***, and ***H.ALL*** follow the same form as those using ***H.VEG*** shown here.

Figure 1.7: Within breeding cycle/across managed treatments prediction accuracies, with and without correction for DTHD

A. Without correction for DTHD



B. With correction for DTHD

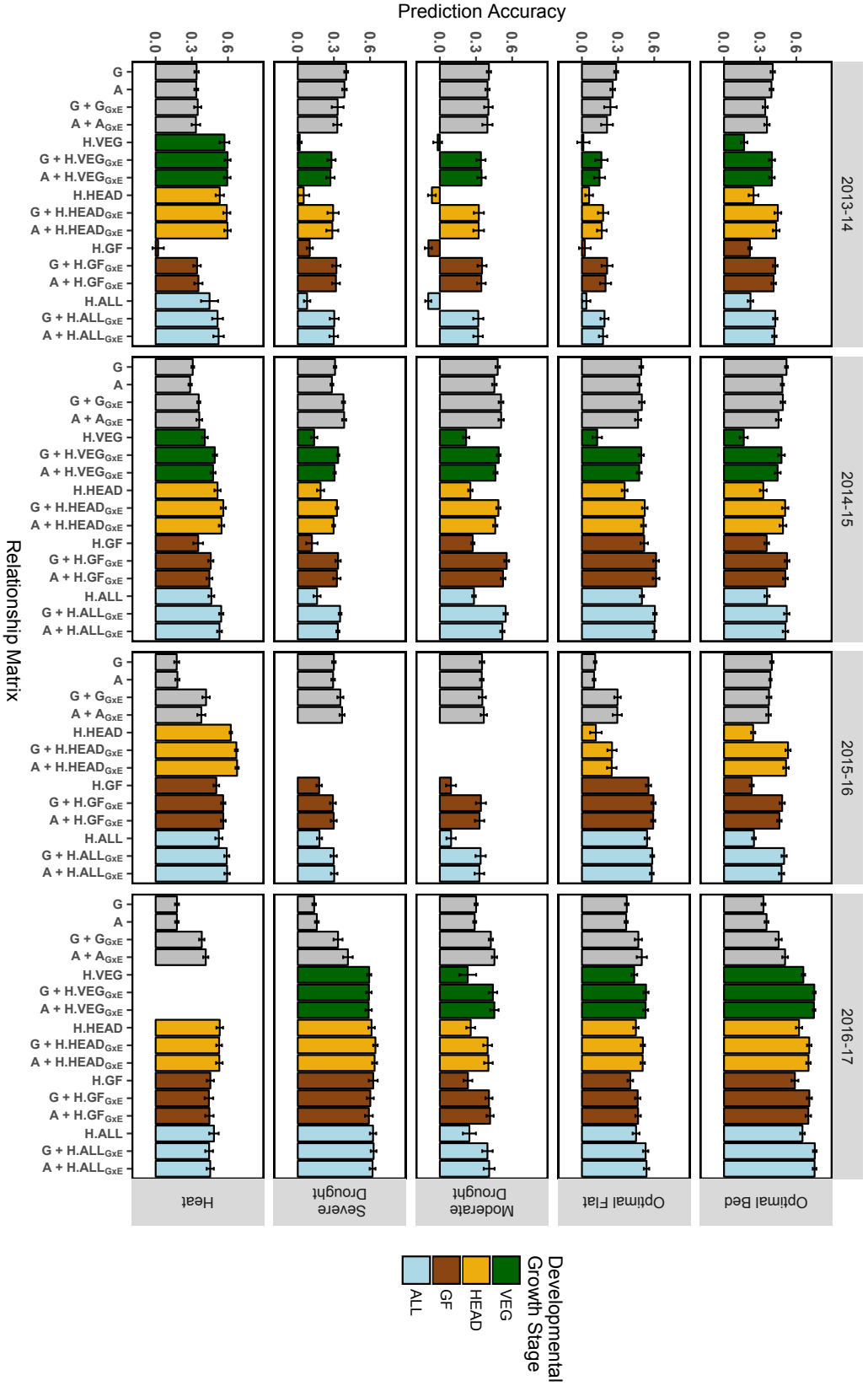


Figure 1.7 (Continued)

Accuracy is expressed as the average Pearson's correlation between predictions and observed BLUPs for GY across 20 random TRN-TST partitions. In each partition, the TRN set consisted of all records from four out of the five managed treatments within the breeding cycle plus 20 percent of records from the fifth managed treatment. Single-kernel models tested were genetic main effects only (**G** or **A**) and hyperspectral reflectance main effects (**H.VEG, H.HEAD, H.GF, H.ALL**). Multi-kernel models assessed were genetic main effects plus genetic GxE (**G + G_{GxE}**, **A + A_{GxE}**) and genetic main effects plus hyperspectral reflectance GxE (e.g., **G + $H.VEG_{GxE}$**, **A + $H.ALL_{GxE}$**, etc.). The color corresponds to the developmental growth stage of the site-year at the time of phenotyping. Error bars are the standard deviation of prediction accuracy for the 20 random partitions.

had lower heritabilities than all other site-years. To summarize trends for the remaining 19 site-years, the results from the single-kernel hyperspectral main effects models for 2013-14 Optimal Flat were removed from the analysis. The accuracies of ***H.VEG*** averaged 0.37 across the site-years where it was tested, which was similar to the accuracies of ***G*** (0.35) and ***A*** (0.34) (Figure 1.7A). ***H.HEAD***, ***H.GF***, and ***H.ALL*** showed slightly higher averages accuracies of 0.44, 0.44, and 0.46, respectively. The greatest differences in prediction accuracy between the hyperspectral reflectance-based models and ***G*** and ***A*** were observed in the 2015-16 and 2016-17 breeding cycles. In these cycles, the ***H.HEAD***, ***H.GF***, and ***H.ALL*** showed accuracies that were on average 0.22 greater than the **G** and **A**. When correcting all models for DTHD, mean accuracies of ***H.VEG*** (0.29), ***H.HEAD*** (0.34), ***H.GF*** (0.30), and ***H.ALL*** (0.34) models were more similar to ***G*** (0.33) and ***A*** (0.32) (Figure 1.7B).

Expanding the single-kernel ***G*** and ***A*** models to account for the $G\times E$ interactions with the $\boldsymbol{G} + \boldsymbol{G}_{\boldsymbol{GxE}}$ and $\boldsymbol{A} + \boldsymbol{A}_{\boldsymbol{GxE}}$ models improved prediction accuracies to a level of 0.43 on average, an increase of 0.08 and 0.09 over the ***G*** and ***A*** models, respectively. These improvements in accuracy were more pronounced during the 2016-17 breeding cycle, recording gains in accuracy of 0.16 and 0.19 for $\boldsymbol{G} + \boldsymbol{G}_{\boldsymbol{GxE}}$ and $\boldsymbol{A} + \boldsymbol{A}_{\boldsymbol{GxE}}$, respectively. Likewise, slightly greater improvements were observed for the Severe Drought (gains of 0.13 and 0.16 for $\boldsymbol{G} + \boldsymbol{G}_{\boldsymbol{GxE}}$ and $\boldsymbol{A} + \boldsymbol{A}_{\boldsymbol{GxE}}$, respectively) managed treatment. Similar trends were observed when correcting for DTHD, though accuracies were 0.04 lower on average.

The multi-kernel models that estimated the main effects using relationship matrices ***G*** or ***A*** and the $G\times E$ interactions using hyperspectral reflectance-derived

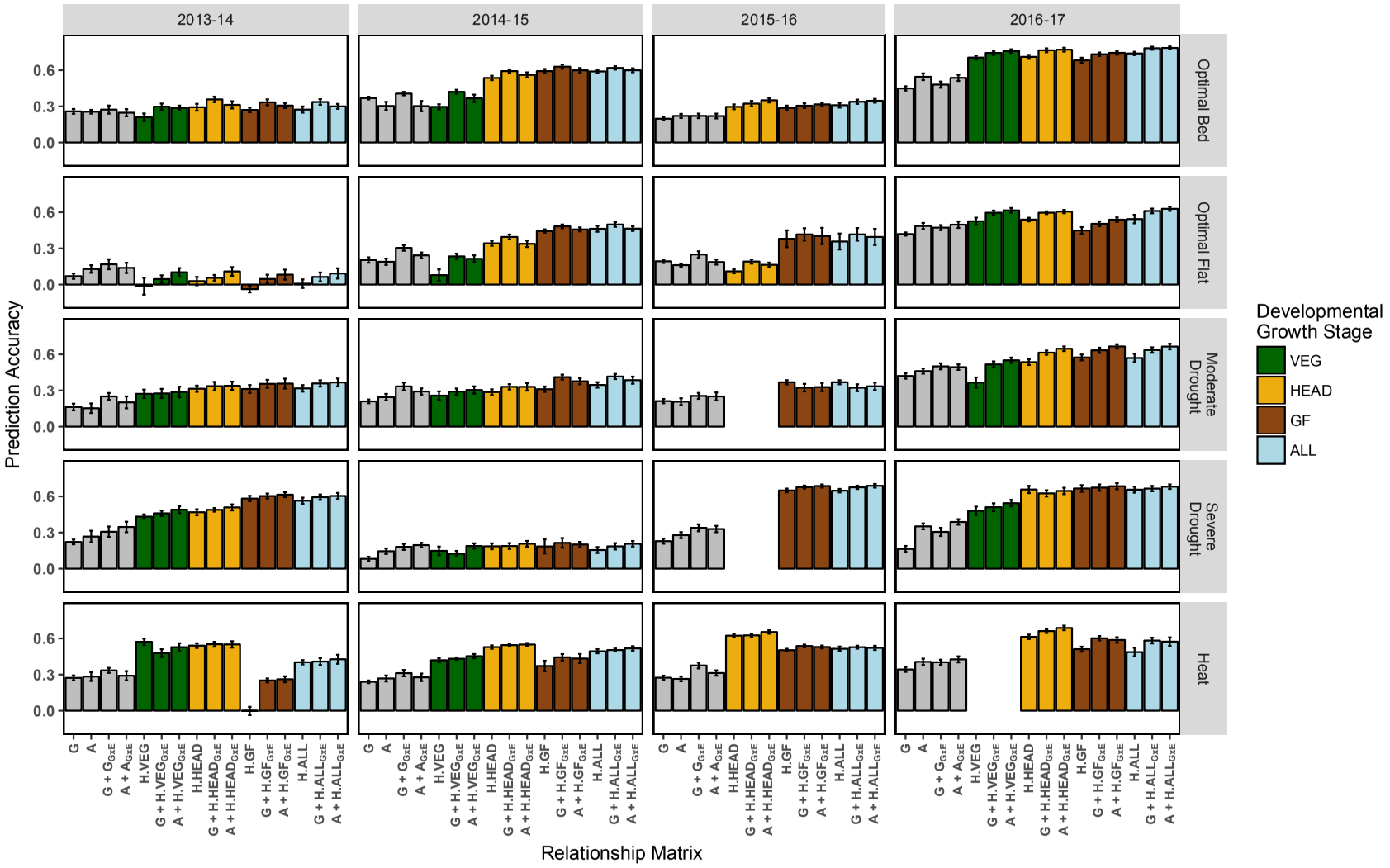
relationship matrices gave the highest accuracies overall. Hyperspectral reflectance-derived relationship matrices ***H.VEG***, ***H.HEAD***, ***H.GF***, and ***H.ALL*** had average accuracies of 0.54, 0.56, 0.56, and 0.58, respectively, when integrated with relationship matrix ***G*** and average accuracies of 0.53, 0.55, 0.55, and 0.57, respectively, when integrated with relationship matrix ***A***. In most site-years, no clear "best model" could be identified among the multi-kernel models within a site-year. Accuracies were similar regardless of during which growth stage the hyperspectral reflectance data were recorded or whether genomic markers or pedigrees were used to model the main effects.

When compared to the $\boldsymbol{G} + \boldsymbol{G}_{\boldsymbol{GxE}}$ and $\boldsymbol{A} + \boldsymbol{A}_{\boldsymbol{GxE}}$ models, the use of hyperspectral reflectance to estimate the $G \times E$ interactions increased prediction accuracies by an average of 0.12 and 0.10, respectively, and by an average of 0.08 and 0.06, respectively, after correcting for DTHD. Compared to the corresponding single-kernel hyperspectral reflectance models, the integration of markers or pedigrees with hyperspectral reflectance in multi-kernel models increased prediction accuracies on an average by 0.11 and 0.17, respectively, and by 0.14 and 0.19, respectively when correcting for DTHD.

Across Breeding Cycles/Within Managed Treatment: The models tested for prediction across breeding cycles/within managed treatment were the same as those tested for prediction within breeding cycle/across managed treatments (Table 1.3). Prediction accuracies are shown in Figure 1.8. On average, accuracies were 0.08 lower in this prediction scheme than for prediction within a breeding cycle and across managed treatments.

Figure 1.8: Across breeding cycles/within managed treatment prediction accuracies, with and without correction for DTHD

A. Without correction for DTHD



B. Without correction for DTHD

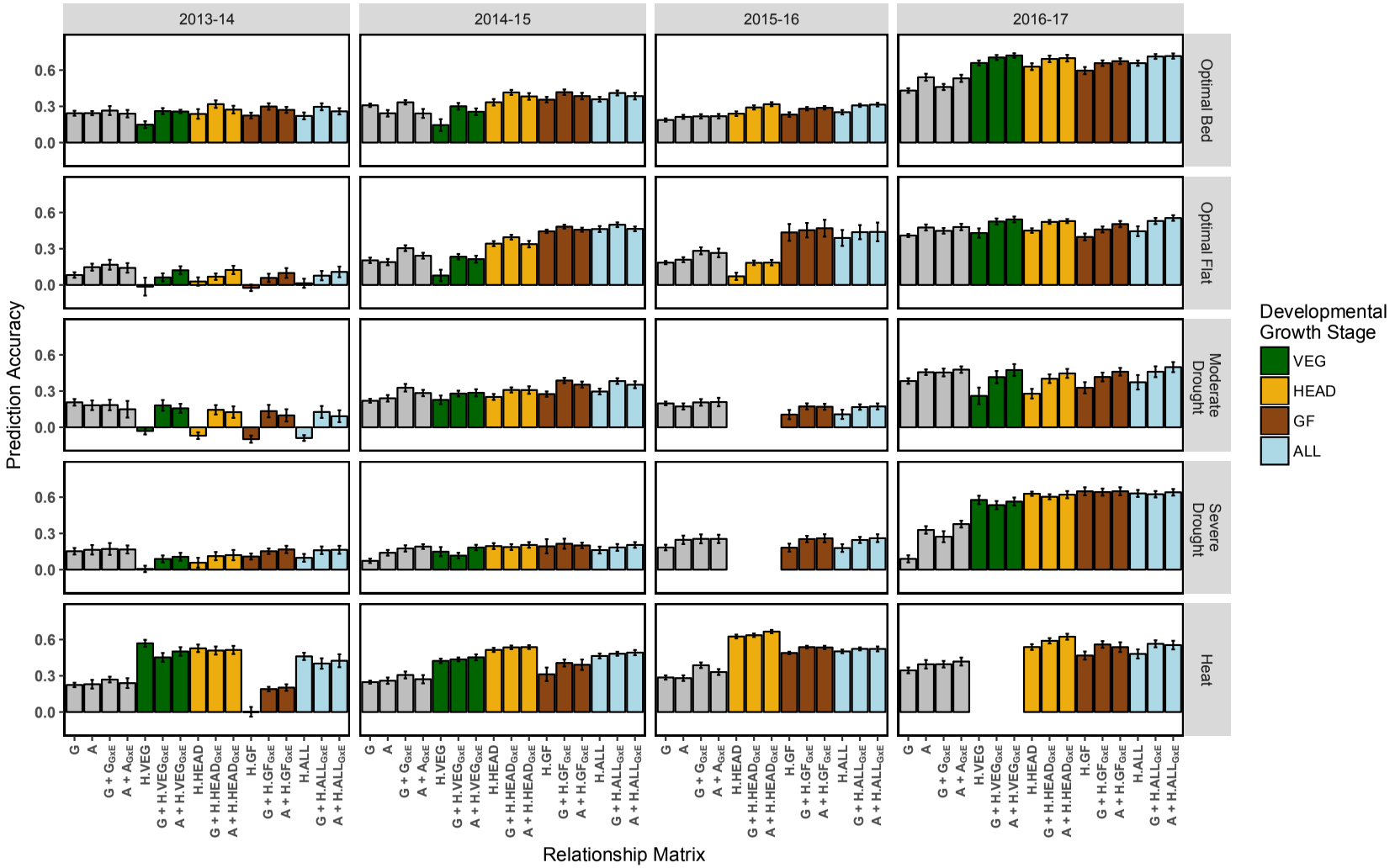


Figure 1.8 (Continued)

Accuracy is expressed as the average Pearson's correlation between predictions and observed BLUPs for GY across 20 random TRN-TST partitions. In each partition, the TRN set consisted of all records from three out of the four of the breeding cycles for the managed treatment plus 20 percent of records from the fourth breeding cycle. Single-kernel models tested were genetic main effects only (**G** or **A**) and hyperspectral reflectance main effects (**H.VEG, H.HEAD, H.GF, H.ALL**). Multi-kernel models assessed were genetic main effects plus genetic GxE (**G + G_{GxE}**, **A + A_{GxE}**) and genetic main effects plus hyperspectral reflectance GxE (e.g., **G + $H.VEG_{GxE}$**, **A + $H.ALL_{GxE}$**, etc.). The color corresponds to the developmental growth stage of the site-year at the time of phenotyping. Error bars are the standard deviation of prediction accuracy for the 20 random partitions.

For the single-kernel hyperspectral main effects models, ***H.GF*** in 2013-14 Heat and all single-kernel ***H*** models for 2013-14 Optimal Flat showed prediction accuracies close to zero. The 2013-14 Heat site-year was phenotyped once during the GF late in the growing season (21 May), which was close to maturity. Figure 4 shows that the correlation between hyperspectral reflectance and GY for that time-point is close to zero for each of the 62 wavelengths. To consider average trends, these were removed from the analysis. Mean accuracies of the ***H.VEG***, ***H.HEAD***, ***H.GF***, and ***H.ALL*** models were 0.37, 0.45, 0.45, and 0.46, respectively (Figure 1.8A). The accuracies of the ***G*** and ***A*** models were slightly lower, recording averages of 0.25 and 0.28, respectively. When correcting for DTHD, the mean accuracies of ***H.VEG***, ***H.HEAD***, ***H.GF***, and ***H.ALL*** models were 0.28, 0.34, 0.32, and 0.34, respectively (Figure 1.8B). While closer in performance to ***G*** and ***A***, the single-kernel hyperspectral main effects models still conferred an advantage in some site-years.

Expanding the ***G*** and ***A*** models to account for the $G\times E$ interactions marginally increased accuracies to 0.32 and 0.31 for the $\boldsymbol{G} + \boldsymbol{G}_{\boldsymbol{GxE}}$ and $\boldsymbol{A} + \boldsymbol{A}_{\boldsymbol{GxE}}$ models, respectively. As with prediction within breeding cycle/across managed treatments, the highest accuracies were observed for the multi-kernel models that estimated the genetic main effects using markers or pedigrees and the $G\times E$ interactions using hyperspectral reflectance. Hyperspectral reflectance matrix ***H.VEG*** had an average accuracy of 0.39 when integrated with ***G*** and 0.41 when integrated with ***A***. ***H.HEAD***, ***H.GF***, and ***H.ALL*** had average accuracies of 0.46, 0.46, and 0.48, respectively, when integrated with ***G*** or ***A***. As before, for most site-years, these multi-kernel models were observed to have similar levels of accuracy within each site-year, irrespective of which

hyperspectral reflectance-derived relationship matrix was used to model the $G \times E$ interactions or whether markers or pedigrees were used to model the genetic main effects.

When compared to the $\boldsymbol{G} + \boldsymbol{G}_{\boldsymbol{GxE}}$ and $\boldsymbol{A} + \boldsymbol{A}_{\boldsymbol{GxE}}$ models, estimating the $G \times E$ interactions using hyperspectral reflectance increased prediction accuracy by an average of 0.12 and 0.14, respectively, and by an average of 0.07 and 0.08, respectively, when correcting for DTHD. Average improvements over the corresponding single-kernel hyperspectral reflectance main effects models ranged from 0.04 to 0.06 and from 0.06 to 0.08 after correcting for DTHD.

Discussion

We proposed a multi-kernel, across-environment GBLUP model that uses relationship matrices derived from genomic markers, pedigrees, and aerial hyperspectral reflectance phenotypes to estimate the genetic main effects and the $G \times E$ interactions in the context of a wheat breeding program. Our study found that deriving a relationship matrix from high-dimensional hyperspectral reflectance phenotypes - as if they were genomic markers - for use in GBLUP can be an effective approach for predicting GY in wheat and in many situations resulted in predictions accuracies that were equivalent or superior to the use of genomic markers or pedigrees. This is consistent with similar studies that used linear and non-linear modeling approaches including ordinary least squares, partial least squares, Bayesian shrinkage, and functional regression methods to predict grain in maize and wheat with hyperspectral reflectance data (Aguate et al., 2017; Montesinos-Lopéz et al., 2017a).

The heritabilities of the hyperspectral wavelengths and their correlations with GY were not homogeneous across time-points, growth stages, breeding cycles, or managed treatments. Wavelengths between 398 and 425 nm were observed to have the lowest heritabilities while those below 500 nm tended to have the weakest correlations with GY. These results are consistent with a similar study of canopy reflectance in wheat (Hansen et al., 2003) that found low signal-to-noise between 400 and 438 nm. According to Mahlein et al. (2015), the hyperspectral imaging systems most commonly used in agriculture have poor sensitivity to reflectance in the blue region of the light spectrum (400-500 nm). Despite this, Montesinos-Lopéz et al. (2017a) showed that removing wavelengths with low heritability did not improve prediction accuracies when using hyperspectral reflectance to predict GY in wheat. Based on this result, we used all hyperspectral wavelengths to build relationship matrices for prediction. However, the blue region of the light spectrum contains important information on the optical properties of plants, including the absorbance maxima of chlorophyll a, chlorophyll b, and carotenoids (Lichenthaler, 1987; Horton et al., 1996). Further advancements in the ability of hyperspectral imaging to accurately record reflectance in the blue region may potentially improve prediction accuracies for grain yield.

When predicting line performance within a site-year of interest, most site-years recorded at least one individual hyperspectral time-point with prediction accuracy equivalent or superior to prediction with markers or pedigrees, suggesting that hyperspectral reflectance phenotypes may be a useful alternative for generating predictions when markers and pedigree are not available. We also found that

combining multiple time-points into hyperspectral BLUEs for a given growth stage generally did not improve accuracy beyond the average accuracy across the individual time-points. In practice, however, GY data would not be available to inform the identification of the most predictive time-points. Therefore, phenotyping at multiple time-points throughout the growing season to develop multi-time-point BLUEs may be an effective approach to buffer against time-points with low prediction accuracy.

It is also possible that some information pertaining to plant growth and development is lost when integrating multiple time-points using means-based approaches. An alternative that may more fully capture temporal variations could be to model a growth curve based on reflectance data for each line in each site-year, as in Verhulst et al. (2011). A greater number of time-points than were measured in this study may be required to accurately develop growth curves. In addition, statistical methods that can integrate multiple curves from a range of hyperspectral bands may be needed. Further research in this area should be performed to compare prediction accuracies of hyperspectral reflectance-based growth curves versus those achieved in this study.

Correcting results for DTHD reduced accuracies when hyperspectral reflectance-derived relationship matrices were used, but not for marker or pedigree matrices. These results suggest that the hyperspectral reflectance measurements are also capturing physiological parameters associated with relative maturity. The greatest reductions in accuracy were observed in the Drought and Severe Drought managed treatments, where GY is typically associated with earliness. This is consistent with Rutkoski et al. (2014), which found that correcting for DTHD in wheat GY reduced

the prediction accuracies of multivariate GS models integrating NDVI and canopy temperature measurements with markers and pedigrees. The CIMMYT bread wheat breeding program maintains a high level of diversity for DTHD in germplasm development due to its wide target of geographic regions. For breeding programs with high levels of variation for DTHD, it may be advisable to perform a correction for DTHD when predicting GY using hyperspectral reflectance so as to avoid indirect selection on relative maturity.

To test our proposed prediction approaches in a multi-environment context, we developed two prediction schemes. In the first, predictions were performed within a breeding cycle for genotypes that have been evaluated under some treatments but not others. In the second, predictions were performed across breeding cycles onto genotypes that were not previously evaluated. Overall, prediction accuracies in the second scheme were lower than for the corresponding models in the first, which is consistent with previous studies showing that predicting the performance of newly developed lines is more challenging than the prediction of lines that have been evaluated in correlated environments (Burgueño et al., 2012; Crossa et al., 2014).

If markers or pedigrees are not available, our results showed that predicting GY with hyperspectral reflectance-derived relationship matrices alone in single-kernel models could provide similar results in terms of accuracy. It should be noted that this study was conducted at the yield trial stage of the breeding program when families contain fewer full-sibs and the variance due to Mendelian sampling is low. As pedigrees do not account for Mendelian sampling, it is possible that prediction with hyperspectral reflectance may be more advantageous than pedigree-based GS at earlier

stages of the breeding program in which families contain greater numbers of full-sibs, though further research is needed to assess the ability of hyperspectral reflectance to distinguish within-family variation.

The optimal prediction accuracies were achieved by building combined models that used markers or pedigrees to model the genetic main effects and hyperspectral reflectance to model the $G \times E$ interactions. However, in prediction across breeding cycles within a managed treatment, the improvements in accuracy with the addition of markers or pedigrees were marginal when compared to the corresponding single-kernel hyperspectral reflectance models. This result was similar to Montesinos-Lopéz et al. (2017b), which found that the addition of markers or pedigrees did not significantly improve accuracy. The modest increases observed when predicting across breeding cycles within a managed treatment may not justify the added investment of marker genotyping.

When considering the optimal developmental growth stages during which to collect hyperspectral reflectance data, there were some differences in accuracy among growth stages when predicting with single hyperspectral phenotyping time-points within a site-year. Overall, the time-points from the HEAD and GF stages showed slightly higher accuracies than those collected during the VEG stage. However, when predicting across site-years, the advantage of phenotyping later in the season was reduced, particularly for the multi-kernel models that combined markers or pedigrees with hyperspectral reflectance-derived relationship matrices, where most often there was no considerable difference in accuracy among the different growth stages. Montesinos-Lopéz et al. (2017a) observed clear optimal time-points for hyperspectral

phenotyping in a multi-environment context. However, their study represented a balanced scenario in which all lines in all environments were phenotyped an equal number of times, and individual phenotyping time-points could be used to predict across environments. While classifying time-points according to the predominant growth stage at the time of phenotyping represents a simple and efficient method for predicting across site-years, it is somewhat difficult to compare results between growth stages. The numbers of time-points observed for each growth stage were not consistent within and across site-years. In addition, the growth stage classifications used were dependent on the predominant growth stage of the site-year and did not reflect the variation in phenology within the site-year at the time of phenotyping. These challenges may warrant further investigation into multi-environment prediction when HTP datasets are unbalanced in the number of time-points observed and at which stage of crop development those observations were recorded.

While our results suggest that hyperspectral reflectance data have the potential to add value to a breeding program by providing accurate GY predictions, the data used in this study were collected on large plot sizes that are suitable for measuring GY *per se*. While our approach may provide breeding programs with GY predictions earlier in the growing season, enabling more efficient allocation of resources at harvest, a potentially greater benefit could come from utilizing this method for smaller plot sizes where there is limited seed available to replicate and reliably assess yield. To address this, we are currently evaluating the use of aerial HTP at the early generation stage when measuring GY is not feasible.

The integration of highly dimensional data, such as hyperspectral reflectance, into relationship matrices for use in GBLUP has several advantages. First, there is no additional development in statistical software required to implement the proposed models. Many options for fitting GBLUP such as "rrBLUP", "GAPIT", and "BGLR" are currently available and are being increasingly used. The models proposed here can be readily implemented in these or other existing software without requiring additional user training. In addition, with the development of improved hyperspectral sensors that register reflectance at wavelengths ranging from 400 to 2500 nm, it is likely that the number of data points available for prediction will continue to increase. By using the relationship matrix-based approach to GBLUP that we have proposed, these increases in data dimensionality will neither influence the complexity of the GBLUP model, nor will they increase the computation time of the model itself. This approach could also be useful for integrating different types of highly dimensional phenotypes for prediction, such as ionomics and metabolomics data (Riedelsheimer et al., 2012). Rincent et al. (2018) recently showed that relationship matrices derived from near infrared spectroscopy absorbance of winter wheat grains and powdered leaf tissue between 400 and 2500 nm provided yield prediction accuracies that were superior to marker-based GBLUP approaches. Further research should be performed to assess the potential for this relationship matrix-based approach to be applied to other forms of highly dimensional biological data.

Conclusion

In this study, we have proposed a multi-kernel GBLUP model that uses genomic marker-, pedigree-, and hyperspectral reflectance-derived relationships

matrices to model the genetic main effects and $G \times E$ interactions across environments within a bread wheat breeding program. We have shown that deriving relationship matrices from aerial hyperspectral reflectance phenotypes can effectively predict GY in wheat within and across managed treatments and breeding cycles. Accuracies when testing single-kernel models using hyperspectral reflectance data alone are similar to those achieved with markers or pedigrees. Our results also show that combining markers/pedigrees with hyperspectral reflectance data in multi-kernel models can increase accuracies over single-kernel approaches, but in some prediction scenarios, these increases were modest. We also suggested a method for addressing the issue of unbalanced HTP datasets involving the classification of time-points according to the predominant developmental growth stage observed at the time of phenotyping. Accuracies of multi-kernel models were roughly equivalent irrespective of the growth stage in which hyperspectral phenotyping was performed. Further research on how best to leverage multi-temporal phenotypes when the amount of data differs across site-years is needed. The methods we have proposed provide a simple and computationally efficient approach for integrating highly dimensional aerial HTP information into genomic selection and should be tested on other forms of high dimensional biological data.

REFERENCES

Aguate, F. M., S. Trachsel, L. González-Pérez, J. Burgueño, J. Crossa, *et al.*, 2017 Use of hyperspectral image data outperforms vegetation indices in prediction of maize yield. Crop Sci. 57: 2517-2524.

Andrade-Sanchez, P., M. A. Gore, J. T. Heun, K. R. Thorp, A. E. Carmo-Silva, *et al.*, 2014 Development and evaluation of a field-based high-throughput phenotyping platform. Funct. Plant Biol. 41: 68-79.

Aparicio, N., D. Villegas, J. Casadesus, J. L. Araus, and C. Royo, 1999 Spectral vegetation indices as nondestructive tools for determining durum wheat yield. Agron. J. 92: 83-91.

Araus, J. L., and J. E. Cairns, 2014 Field high-throughput phenotyping: the new crop breeding frontier. Trends Plant Sci.19: 52-61.

Bernardo, R., 2010 *Breeding for quantitative traits in plants*. Stemma Press, Woodbury, Minnesota.

Burgueño, J., G. de los Campos, K. Weigel, and J. Crossa, 2012 Genomic prediction of breeding values when modeling genotype × environment interaction using pedigree and dense molecular markers. Crop Sci. 52: 707–719.

Cabrera-Bosquet, L., J. Crossa, J. von Zitzewitz, M. D. Serret, and J. L. Araus, 2012 High-throughput phenotyping and genomic selection: the frontiers of crop breeding coverage. J. Integr. Plant Biol. 54: 312-320.

Crain, J., Y. Wei, J. Barker, S. M. Thompson, P. D. Alderman, *et al.*, 2016 Development and deployment of a portable field phenotyping platform. Crop Sci. 56: 965-975.

Crain, J., S. Mondal, J. E. Rutkoski, R. Singh, and J. Poland, 2018 Combining high-throughput phenotyping and genomic information to increase prediction and selection accuracy in wheat breeding. Plant Genome 11: 170043.

Crossa, J., P. Pérez, J. Hickey, J. Burgueño, L. Ornella, *et al.*, 2014 Genomic prediction in CIMMYT maize and wheat breeding programs. Heredity 112: 48-60.

Crossa, J., P. Pérez-Rodríguez, J. Cuevas, O. Montesinos-López, D. Jarquín, *et al.*, 2017 Genomic selection in plant breeding: methods, models, and perspectives. Trends Plant Sci. 22: 961-975.

Daetwyler, H. D., M. P. L. Calus, R. Pong-Wong, G. de los Campos, and J. M. Hickey, 2013 Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. Genetics 193:347-365.

de los Campos, G., and P. Pérez-Rodríguez, 2014 Bayesian generalized linear regression. R package version 1.0.5. URL https://CRAN.R-project.org/package=BGLR.

Elshire, R. J., J. C. Glaubitz, Q. Sun, J. A. Poland, K. Kawamoto, *et al.*, 2011 A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6: e19379.

Endelman, J. B., 2011 Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Genome 4: 250-255.

Endelman, J. B., and J. L. Jannink, 2012 Shrinkage estimation of the realized relationship matrix. G3 Genes Genom. Genet. 2: 1405-1413.

Furbank, R. T., and M. Tester, 2011 Phenomics – technologies to relieve the phenotyping bottleneck. Trends Plant Sci. 16: 635-644.

Gilmour, A. R., B. J. Gogel, B. R. Cullis, S. J. Welham, and R. Thompson, 2014 ASReml user guide release 4.1 structural specification. VSN International, Hemel Hempstead, UK.

Govaerts, Y. M., M. M. Verstraete, B. Pinty, and N. Gobron, 1999 Designing optimal spectral indices: a feasibility and proof of concept study. Int. J. Remote Sens. 1999: 1853-1873.

Gueymard, C., 1995 *A simple model of the atmospheric radiative transfer of sunshine: algorithms and performance assessment.* Florida Solar Energy Center, Cocoa, FL, USA.

Gueymard, C. A., 2005 Interdisciplinary applications of a versatile spectral solar irradiance model: a review. Energy: 30, 1551-1576.

Haghighattalab, A., L. González-Pérez, S. Mondal, D. Singh, D. Schinstock, *et al.*, 2016 Application of unmanned aerial systems for high throughput phenotyping of large wheat breeding nurseries. Plant Methods: 12, 35.

Hansen, P. M., and J. K. Schjoerring, 2003 Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. Remote Sens. Environ. 86:542-553.

Heslot, N., J. Jannink, and M. E. Sorrells, 2013 Using genomic prediction to characterize environments and optimize prediction accuracy in applied breeding data. Crop Sci. 53: 921-933.

Horton, P., A. V. Ruban, and R. G. Walters, 1996 Regulation of light harvesting in green plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. 47:655-684.

Jarquín, D., J. Crossa, X. Lacaze, P. Du Cheyron, J. Daucourt, *et al.*, 2014 A reaction norm model for genomic selection using high-dimensional genomic and environmental data. Theor. Appl. Genet. 127: 595-607.

Labus, M., G. Nielsen, R. Lawrence, R. Engel, and D. Long, 2002 Wheat yield estimates using multi-temporal NDVI satellite imagery. Int. J. Remote Sens. 23: 4169-4180.

Li, L., Q. Zhang, and D. Huang, 2014 A review of imaging techniques for plant phenotyping. Sensors 14: 20078-20111.

Lichtenthaler, H. K., 1987 Chlorophylls and carotenoids: pigments of photosynthetic biomembranes. Methods Enzymol. 148:350-382.

López-Cruz, M., J. Crossa, D. Bonnett, S. Dreisigacker, J. Poland, *et al.*, 2015 Increased prediction accuracy in wheat breeding trials using markers × environment interaction genomic selection model. G3 Genes Genom. Genet. 5: 569-582.

Mahlein, A.K., S. Hammersley, E. C. Oerke, H. W. Dehne, H. Goldbach, *et al.*, 2015 Supplemental blue LED lighting array to improve signal quality in hyperspectral imaging of plants. Sensors (Basel) 15:12834-12840.

McLaren, C. G., L. Ramos, C. Lopez, and W. Eusebio, 2000 Applications of the genealogy management system, pp. 5.8-5.13 in *International Crop Information System. Technical Development Manual, version VI*, edited by C. G. McLaren, J.W. White and P.N. Fox. CIMMyT, México: CIMMyT and IRRI.

Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard, 2001 Prediction of total genetic value using genome-wide dense marker maps. Genetics 157: 1819-1829.

Montesinos-López, O. A., A. Montesinos-López, J. Crossa, G. de los Campos, G. Alvarado, *et al.*, 2017a Predicting grain yield using hyperspectral reflectance in wheat breeding data. Plant Methods 13: 4.

Montesinos-López, A., O. A. Montesinos-López, J. Cuevas, W. A. Mata-López, J. Burgueño, *et al.*, 2017b Genomic Bayesian functional regression models with interactions for predicting wheat grain yield using hyperspectral reflectance. Plant Methods 13: 62.

Pauli, D., S. C. Chapman, R. Bart, C. N. Topp, C. J. Lawrence-Dill, *et al.,* 2016 The quest for understanding phenotypic variation via integrated approaches in the field environment. Plant Physiol. 172: 622-634.

Pérez-Rodríguez, P., J. Crossa, J. Rutkoski, J. Poland, R. Singh, *et al.*, 2017 Single-step genomic and pedigree genotype × environment interaction models for predicting wheat lines in international environments. Plant Genome 10. doi:10.3835/plantgenome2016.09.0089.

Poland, J., J. Endelman, J. Dawson, J. Rutkoski, S. Wu, *et al*., 2012 Genomic selection in wheat breeding using genotyping-by-sequencing. Plant Genome 5: 103-113.

R Core Team, 2018 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Riedelsheimer, C., A. Czedik-Eysenberg, C. Grieder, J. Lisec, F. Technow, *et al.*, 2012 Genomic and metabolic prediction of complex heterotic traits in hybrid maize. Nat. Genet. 44: 217-220.

Rincent, R., J. P. Charpentier, P. Faivre-Rampant, E. Paux, J. Le Gouis, *et al.*, 2018 Phenomic selection is a low-cost and high-throughput method based on indirect predictions: proof of concept on wheat and poplar. G3 Genes Genom. Genet. 8 doi: 10.1534/g3.118.200760.

Rodrigues, F., G. Blasch, P. Defourny, J. Ortiz-Monasterio, U. Schulthess, *et al.*, 2018 Multi-temporal and spectral analysis of high-resolution hyperspectral airborne imagery for precision agriculture: assessment of wheat grain yield and protein content. Remote Sens. 10: 930.

Rutkoski, J. E., J. A. Poland, S. Mondal, E. Autrique, L. González-Pérez, *et al.*, 2016 Canopy temperature and vegetation indices from high-throughput phenotyping improve accuracy of pedigree and genomic selection for grain yield in wheat. G3 Genes Genom. Genet. 6: 2799-2808.

Sun, J., J. E. Rutkoski, J. A. Poland, J. Crossa, J. Jannink, *et al*., 2017 Multitrait, random regression, or simple repeatability model in high-throughput phenotyping data improve genomic prediction for wheat grain yield. Plant Genome 10: 1-12.

Verhulst, N., B. Govaerts, V. Nelissen, K. D. Sayre, J. Crossa, *et al.*, 2011 The effect of tillage, crop rotation and residue management on maize and wheat growth and development evaluated with an optical sensor. Field Crops Res. 120: 58-67.

Viña, A., A. A. Gitelson, D. C. Rundquist, G. Keydan, B. Leavitt, *et al.*, 2003 Monitoring maize (*Zea mays* L.) phenology with remote sensing. Agron. J. 96: 1139-1147.

Viña, A., A. A. Gitelson, A. L. Nguy-Robertson, and Y. Peng, 2011 Comparison of different vegetation indices for the remote assessments of green leaf area index of crops. Remote Sens. Environ. 115: 3468-3478.

White, J.W., P. Andrade-Sanchez, M. A. Gore, K. F. Bronson, T. A. Coffelt, *et al.*, 2012 Field-based phenomics for plant genetics research. Field Crops Res. 133: 101-112.

Zhang, X., M. A. Friedl, C. B. Schaaf, A. H. Strahler, J. C. F. Hodges, *et al*., 2003 Monitoring vegetation phenology using MODIS. Remote Sens. Environ. 84: 471-475.

Table S1.1: Prediction strategies used for assessing model accuracy

A. WITHIN SITE-YEAR

	2013-14		2014-15	2015-16	2016-17
Optimal Bed	TRN	TST			
Optimal Flat					
Moderate Drought					
Severe Drought					
Heat					

B. WITHIN BREEDING CYCLE/ACROSS MANAGED TREATMENTS

	2013-14		2014-15	2015-16	2016-17
Optimal Bed	TST	TRN			
Optimal Flat	TRN				
Moderate Drought	TRN				
Severe Drought	TRN				
Heat	TRN				

C. ACROSS BREEDING CYCLES/WITHIN MANAGED TREATMENT

	2013-14		2014-15	2015-16	2016-17
Optimal Bed	TST	TRN	TRN	TRN	TRN
Optimal Flat					
Moderate Drought					
Severe Drought					
Heat					

These tables represent an example of how phenotypic data records were partitioned and assigned to training and test sets for the three prediction strategies tested. Each cell represents one of the 20 site-years tested. Records allocated to the training set are indicated by TRN, while records assigned to the test set are indicated by TST. Within site-year prediction was performed considering records within a single site-year only. In this example, 80 percent of the records from the 2013-14 Optimal Bed site-year were assigned to the TRN set, while the remaining 20 percent from the 2013-14 Optimal Bed site-year were assigned to the TST set. Within breeding cycle/across managed treatment prediction was performed within a single breeding

Table S1.1 (Continued)

cycle but across the five managed treatments. In this example, the Optimal Flat, Moderate Drought, Severe Drought, and Heat treatments from the 2013-14 breeding cycle were assigned to the TRT set, in addition to 20 percent of the records from the 2013-14 Optimal Bed site-year. The remaining 80 percent of records from the 2013-14 Optimal Bed site-year were assigned to the TST set. Across breeding cycles/within managed treatment prediction was performed across the four breeding cycles but within a single managed treatment. In this example, the Optimal Bed treatments from the 2014-15, 2015-16, and 2016-17 breeding cycles were assigned to the TRN set, in addition to 20 percent of the records from the 2013-14 Optimal Bed site-year. The remaining 80 percent of records from the 2013-14 Optimal Bed site-year were assigned to the TST set. For prediction of each site-year with each prediction strategy, 20 TRN-TST partitions were implemented.

Table S1.2: Summary statistics of grain yield data for each of the 20 site-years tested

Breeding Cycle	Managed Treatment	Mean (t ha^{-1})	Standard Deviation (t ha^{-1})	Range (t ha^{-1})	Broad Sense Heritability
2013-14	Optimal Bed	6.08	0.51	3.38-7.62	0.71
	Optimal Flat	6.03	0.61	3.82-7.60	0.58
	Moderate Drought	3.68	0.30	2.33-4.64	0.85
	Severe Drought	2.28	0.43	0.29-3.31	0.94
	Heat	2.08	0.46	0.05-3.00	0.93
2014-15	Optimal Bed	5.61	0.48	3.56-6.78	0.82
	Optimal Flat	5.72	0.52	3.74-6.97	0.74
	Moderate Drought	4.52	0.36	2.96-5.69	0.79
	Severe Drought	3.80	0.51	1.05-5.16	0.67
	Heat	2.73	0.35	1.51-4.38	0.84
2015-16	Optimal Bed	7.15	0.37	5.89-8.29	0.75
	Optimal Flat	6.97	0.48	4.68-8.74	0.68
	Moderate Drought	3.27	0.33	0.67-4.04	0.82
	Severe Drought	3.66	0.40	1.88-4.99	0.93
	Heat	1.63	0.46	0.04-2.72	0.79
2016-17	Optimal Bed	6.51	0.55	4.26-7.80	0.71
	Optimal Flat	6.28	0.68	3.09-7.96	0.80
	Moderate Drought	4.84	0.42	3.30-5.97	0.82
	Severe Drought	2.63	0.50	0.50-3.94	0.83
	Heat	3.95	0.44	1.53-5.11	0.88

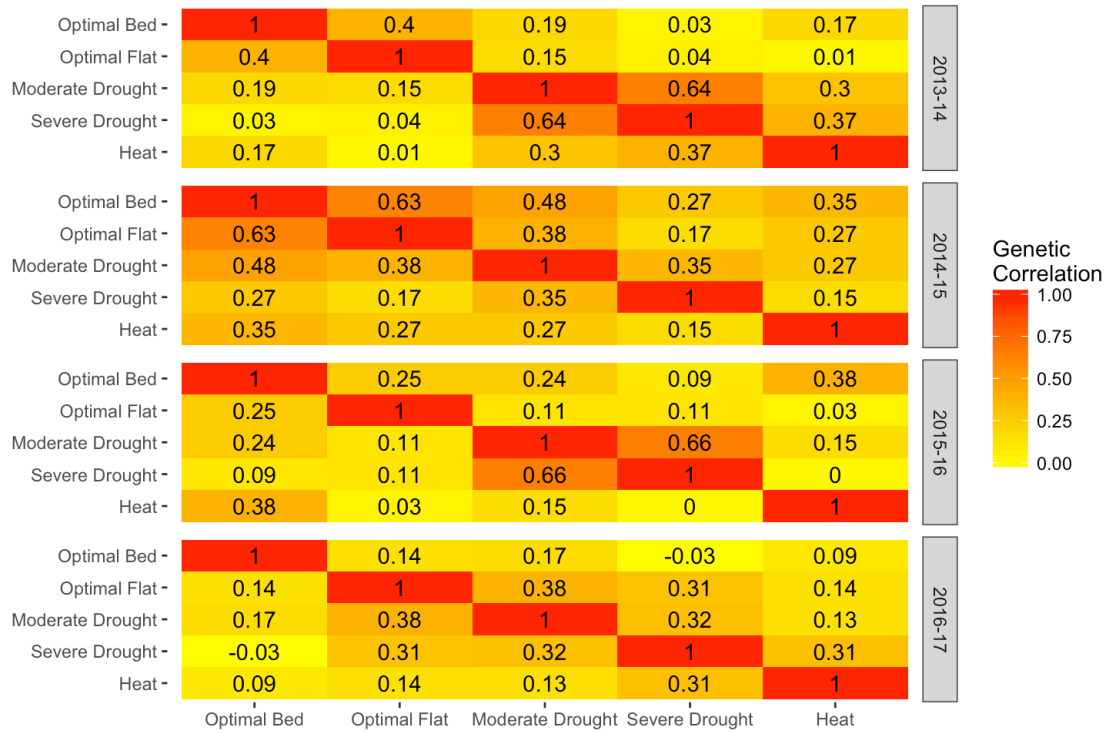


Figure S1.1: Genetic correlations for grain yield between managed treatments within breeding cycles

For each breeding cycle, the genetic correlations for grain yield between managed treatments are reported. Strong and weak correlations are shown in shades of red and yellow, respectively. Intermediate correlations are shaded in orange.

CHAPTER 2

HIGH-THROUGHPUT PHENOTPYING FOR INDIRECT SELECTION ON WHEAT GRAIN YIELD AT THE EARLY-GENERATION SEED-LIMITED STAGE IN BREEDING PROGRAMS

Abstract

Wheat breeding programs require one or more generations of seed increase before replicated yield trials can be sown. Extensive phenotyping at this stage of the breeding cycle is challenging due to the large number of individuals present in the breeding program, and breeders typically rely on visual selection of small, unreplicated seed increase plots for promotion of individuals to replicated yield trials. With the development of aerial high-throughput phenotyping technologies, breeders now have the ability to rapidly phenotype tens of thousands of individuals for traits that may be used for more effective indirect selection for grain yield. We evaluated early generation material in the irrigated bread wheat (*Triticum aestivum* L.) breeding program at the International Maize and Wheat Improvement Center to determine if aerial measurements of vegetation indices assessed on small, unreplicated plots were predictive of grain yield. A population of 2,016 individuals were sown both as replicated yield trials and as small, unreplicated plots during two breeding cycles. Vegetation indices collected with an unmanned aerial vehicle in the small plots were observed to be heritable and more predictive of grain yield than univariate genomic selection, though multi-trait genomic selection approaches that combined genomic

information with the aerial phenotypes were found to have the highest predictive abilities overall. A related experiment showed that selection approaches for grain yield based on vegetation indices could be more effective than visual selection; however, selection on the vegetation indices alone would have also resulted in a directional response in phenology due to confounding between those traits. A restricted selection index was proposed for improving grain yield without affecting the distribution of phenology in the breeding population. The results of these experiments provide a promising outlook for the use of aerial high-throughput phenotyping traits to improve selection at the early generation seed-limited stage of wheat breeding programs.

Introduction

For small grains and some other crops such as soybean and rice, limited seed yield per plant can delay replicated yield testing until sufficient quantities of seed are accumulated through one or more generations of seed increase. Breeding lines are sown in rows or small plots in the field for the purpose of seed increase prior to yield testing. The generation at which this occurs depends on the crop and breeding scheme followed, though in theory this may begin as early as the F_3 with yield testing at the F_4. The ability to accurately cull breeding lines with low potential the early generation seed-limited stage is highly desirable because it reduces subsequent expenditures on costly yield testing (Brennan, 1988).

Selection at the early generation seed-limited stage of a breeding program relies on the assumption that the phenotype of an individual when sown in a small, unreplicated plot is predictive of its eventual performance in a larger replicated yield trial. Selection on small plots of unreplicated material can be applied most effectively

for qualitative traits controlled by a limited number of genetic loci with major effects (Snape and Simpson, 1984). Selection on quantitative traits, however, becomes more challenging due to the high level of influence of non-genetic effects (Bernardo, 2003).

Numerous studies have shown that plants exhibit plastic phenotypic responses to the competitive stress imposed by their neighbors (White and Harper, 1970; Edmeades and Daynard, 1979; Dornbusch et al., 2011). The level of competitiveness occurring in small plots does not mimic that which occurs in yield trials or farmers fields. Therefore; it can be difficult to select for grain yield at the early generation seed-limited stages of breeding programs. Indeed, the yields of single wheat plants or small plots have been observed to exhibit low heritability (Fonseca and Patterson, 1968; Mitchell et al., 1982) and weak correlations with the yields of replicated yield trials (Syme, 1972; Knott and Kumar, 1975). Non-genetic factors can also drive variation in trait values when breeding material is evaluated in small plots (reviewed by Rebetzke et al., 2014). For example, variation in canopy height and architecture can lead to light competitiveness among different breeding lines sown in adjacent plots (Clarke et al., 1998).

A potential strategy to improve selection accruacy at the early generation seed-limited stages of wheat breeding programs would be to perform indirect selection on a highly heritable secondary trait or traits that exhibit high correlation with grain yield when measured in replicated yield trials. Several measurable physiology traits in wheat have been shown to fit these criteria. For example, Rebetzke et al. (2002) demonstrated that selection for reduced carbon isotope discrimination at the early generation stage increased grain yield and aboveground biomass in wheat. Likewise,

significant responses in grain yield were observed in seven out of nine segregating populations of wheat when indirect selection on harvest index was imposed (Borghi et al., 1997). Quail et al. (1989) found significant correlations between wheat grain yield in the F_7/F_8 generation with fruiting efficiency at maturity, or the ratio of the number of grains per spike to the dry weight of the spike, measured at the F_3 generation. While these and other studies demonstrate the potential for indirect selection on secondary traits, the feasibility of applying these approaches within breeding programs remains limited due to the high cost and time required to evaluate these traits in the field on large numbers of lines. For indirect selection at the early generation seed-limited stage to be feasible, the phenotyping must be high-throughput to cover the large number of breeding lines sown and the cost not exceed the amount saved on subsequent yield trials by culling low potential material based on secondary traits.

Recent applications of advances in the remote sensing industry to plant breeding have aimed to enable rapid, field-based phenotyping of physiological and other traits on large numbers of crop plants while reducing labor, time, and cost with respect to traditional approaches (Cabrera-Bosquet et al., 2012; Araus and Cairns 2014; Pauli et al., 2016). In recent years, a range of ground- and aerial-based high-throughput phenotyping (HTP) systems utilizing various cameras and sensors have been developed and tested to expand the efficiency of phenotypic data collection (Andrade-Sanchez et al., 2014; Crain et al., 2016; Haghighattalab et al., 2016; Watanabe et al., 2017). Vegetation indices (VIs) provide an integrated measurement of canopy structure and photosynthetic activity based on the amount of light reflected off of the crop canopy (Huete et al., 2000) and can be routinely measured with HTP

systems (Reynolds and Langridge, 2016). A number of studies have demonstrated strong correlations between VIs and grain yield in wheat (Babar et al., 2006; Prasad et al., 2007; Prasanna et al., 2013). In addition, spectral reflectance traits have been shown to improve the accuracy of grain yield prediction in wheat when integrated with genetic information in genomic selection (GS) (Rutkoski et al., 2016; Sun et al., 2017; Crain et al., 2018; Krause et al., 2019). These previous efforts have leveraged the replicated yield trial stage of breeding programs to evaluate the ability to predict grain yield using traits collected with aerial HTP systems. While these approaches may serve to provide breeders with yield predictions earlier in the season, thereby facilitating greater resource-use efficiency at harvest, the potential cost savings may not justify the expense of collecting HTP traits. A greater benefit could come from applying HTP at the early stages of the breeding program when limited seed availability prevents the reliable assessment of grain yield. If HTP traits can facilitate greater selection accuracy, programs can select at a higher intensity, thereby promoting fewer lines to more expensive replicated yield trials.

Several plant breeding strategies in small grains – including pedigree selection, bulk selection, single seed descent, and doubled haploids, among others – require one or more generations of spaced plantings for seed increase prior to replicated yield testing, opening up the possibility for indirect selection on secondary traits using HTP. The irrigated bread wheat breeding program at the International Maize and Wheat Improvement Center (CIMMYT) utilizes a "selected bulk" scheme between the F_3 and F_5 generations in which one spike from each selected plant is harvested and bulked with 30 to 50 spikes from the same cross to form the next generation (Singh et al.,

1996). At the F_5 or F_6 generation, individual plants or spikes are selected and promoted to the next stage independently, which are sown as double- or triple-row 0.8m × 0.8m unreplicated small plots (SP) at the Campo Experimental Norman E. Borlaug (CENEB) in Ciudad Obregón, Sonora, México. This method allows tens-of-thousands of breeding lines to be planted within a small area for seed increase and visual selection prior to the first year of replicated yield testing. Individuals are selected for uniformity, disease resistance, and other agronomic characteristics and then harvested in bulk. A second round of visual selection is then performed on the seed following harvest to retain individuals with desirable grain characteristics for promotion to the first year of replicated yield trials (YT) (van Ginkel et al., 2002).

Detailed phenotyping of the SP stage of CIMMYT's irrigated bread wheat breeding program is challenging due to the large number of individuals sown each breeding cycle. Aerial-based HTP platforms allow screening of thousands to tens-of-thousands of plots within a short time span and with less labor, opening up the possibility of selection on secondary HTP traits at the SP stage of the CIMMYT wheat breeding program. In addition, because the breeding program also sows YT, which are phenotyped for grain yield, at CENEB simultaneously, models utilizing HTP traits to predict grain yield can be trained with HTP and grain yield records from the YT and then applied to the SP to estimate grain yield predictions during the same breeding cycle. The objective of this study was to assess the potential for utilizing VIs collected with aerial HTP systems at the early generation seed-limited stage of the breeding program to improve selection for grain yield in wheat.

Materials and Methods

Field Trial Design

During the 2016-17 and 2017-18 breeding cycles, two sets of 1,008 entries of CIMMYT's first-year yield trials (YT) were evaluated for agronomic and HTP traits at CENEB in Ciudad Obregón, Mexico. A unique set of 1,008 entries was assessed in each cycle with no overlap in individuals between cycles except for check varieties.

Each year, the 1,008 entries were arranged in 36 trials containing 28 entries each. Trials were sown in an α-lattice design with 2 replicates and 6 incomplete blocks per replicate, each block containing 5 entries. Check varieties "Borlaug100 F2014" and "Kachu #1" were included twice per replicate.

Each set of 1,008 entries contained full-sib families ranging in size from 1 to 34 full-sibs with an average of 4 full-sibs per family. In the first replicate of the YT, entries were ordered by pedigree, such that full-sibs were sown adjacent to one another in families in a serpentine order with the check varieties appearing as the first two plots of the replicate. In the second replicate, entries and checks were completely randomized. Plots in the YT were 2.8m × 1.6m in size and consisted of two beds containing three rows each.

To empirically evaluate the correlation between small plot performance and replicated yield trials, the same 1,008 entries in each breeding cycle were sown in parallel at CENEB as unreplicated 1m × 0.8m small plots (SP) containing three rows each and were evaluated for both agronomic and HTP traits. The 1,008 entries were sown in a serpentine pattern with the order reflecting that of the first replicate of each trial in the YT. During the 2016-17 breeding cycle, the two checks "Borlaug 100

F2014" and "Kachu #1" appeared at the beginning of each set of 28 individuals. In the 2017-18 breeding cycle, the checks were randomly dispersed within each set of 28 individuals. Additional check varieties "MISR 1" and "Baj #1" were added during the 2017-2018 breeding cycle and sown ten times each at random locations within the field.

The YT and SP were sown on 29 November 2016 and 25 November 2016 during the 2016-17 cycle and on 11 November 2017 and 23 November 2017 during the 2017-18 cycle, respectively. Both the YT and SP were visually evaluated during the growing season for days to heading (DTHD) and days to maturity (DTMT). DTHD and DTMT were evaluated as the number of days to reach the heading or maturity stages starting from the date of the first irrigation if sowing was on dry soil or from the sowing date if sowing was in pre-irrigated fields. In the YT, DTHD and DTMT were evaluated on the first replicate only, due to the high heritability of these traits. Although the breeding program does not typically assess grain yield at the SP stage, both the SP and YT were harvested using plot combine harvesters and then weighed manually for grain yield (GY).

In addition to this proof of concept experiment, the SP stage of the formal CIMMYT breeding program was assessed for HTP traits during the 2016-17 breeding cycle. These individuals were sown as 0.8m × 0.8m unreplicated SP containing two rows each. Plots were arranged in a serpentine pattern with full-sibs sown adjacent to one another in families. A check variety was sown after every 50 plots, alternating between varieties "Borlaug 100 F2014" and "Kachu #1." Those individuals that were visually selected by breeders were promoted to first-year YT in 2017-18. Therefore,

the individuals evaluated as YT and SP in 2017-18 as part of the proof of concept experiment had also been evaluated for HTP as part of the SP stage of the breeding program during the previous cycle. These plots are herein referred to as SP_{BP}. The SP_{BP} were also assessed for DTHD and DTMT on a qualitative scale with respect to the nearest "Borlaug 100 F2014" check. Plots were scored as "early", "mid", "late", and "very late" for DTHD and DTMT. The check "Borlaug 100 F2014" was considered to have "mid" DTHD and DTMT. The first 300 plots were harvested and weighed manually for GY, enabling a comparison of response to selection for visual versus HTP selection.

Aerial High-Throughput Phenotyping

Ground Control Points: The trials were prepared for aerial imaging by distributing ground control points (GCPs) within the field. The GPS locations of each GCP were measured with a R4 RTK GPS (Trimble, Sunnyvale, CA) with a horizontal accuracy of 0.025m and a vertical accuracy of 0.035m.

Aerial HTP Equipment: A Matrice 100 quadcopter unmanned aerial vehicle (UAV) (DJI, Shenzhen, China) equipped with a MicaSense RedEdge camera (MicaSense, Seattle, WA) was flown over both the YT and SP at multiple time-points during the grain filling developmental growth stage. The Matrice 100 has a maximum takeoff weight of 3600g and can hover for 22-28min with no payload. Following a modification to hold an additional battery pack, the maximum no payload hover time was increased to 33-40min. The MicaSense RedEdge weighs 150g and captures one image in 12-bit RAW format per second. Reflectance was recorded at five broadband wavelengths: blue (460-510nm), green (545-575nm), red (630-690nm), NIR (820-

860nm), and red edge (712-722nm). During the 2017-2018 breeding cycle, a MicaSense Downwelling Light Sensor (MicaSense, Seattle, WA) integrated with the RedEdge was added on the top of the UAV to improve calibrations when ambient light conditions were changing throughout the flight.

Calibration Panel: Prior to each flight, images of a MicaSense Calibrated Reflectance Panel (MicaSense, Seattle, WA) were captured with the MicaSense RedEdge for use in post-processing of the imagery to account for the ambient light conditions at the time of flight. As provided by MicaSense, the absolute reflectance values of the panel were: 0.366 (blue), 0.366 (green), 0.391 (red), 0.411 (NIR), and 0.399 (red edge).

Acquisition of Aerial Imagery: During the 2016-2017 breeding cycle, the UAV was flown at an altitude of 30m with a speed of 12 km/hr, giving an image resolution of 2 cm/pixel. In the 2017-2018 breeding cycle, the altitude and speed of the UAV were reduced to 16m and 7 km/hr, increasing the image resolution to 1 cm/pixel. Flights were conducted within 1h of solar noon to minimize variation due to the solar zenith angle (Gu et al., 1992).

Aerial Imagery Post-Processing and Data Extraction: A semi-automated data analysis pipeline presented in Haghighattalab et al. (2016) was utilized to analyze each image dataset and extract plot-level phenotypic values. In summary, the following steps were performed to process each HTP data collection time-point: 1) aligning the aerial images and building sparse point clouds, 2) importing GCP GPS coordinates and geo-referencing the images, 3) constructing a dense point cloud, 4) creating a digital elevation model (DEM), 5) generated an orthomosaic with the DEM, 6)

calculating vegetation indices (VIs), and 6) extracting the plot-level data. The VIs measured were: normalized difference vegetation index (NDVI), green NDVI (GNDVI), and red-edge NDVI (RENDVI). Each VI was calculated following:

$$\frac{\rho_X - \rho_{Red}}{\rho_X + \rho_{Red}} \quad (1)$$

where ρ_{Red} represents the spectral reflectance recorded in the red region of the light spectrum and ρ_X represents the spectral reflectance recorded in the near-infrared, green, and red-edge regions for NDVI, GNDVI, and RENDVI, respectively.

Genotyping and Relationship Matrix Development

The 2,016 entries were genotyped with genotyping-by-sequencing (GBS) (Elshire et al., 2011). The Tassel 5 GBS v2 pipeline was used to process raw Illumina reads and call single nucleotide polymorphism (SNP) markers with minor allele frequency (MAF) of 0.01 (Glaubitz et al., 2014). The unique tags from the Tassel 5 GBS v2 pipeline were aligned to the International Wheat Genome Sequencing Consortium's (IWGSC) RefSeq v1.0 assembly of Chinese Spring using Bowtie2 to get a maximum number of tags with unique mapping (Appels et al., 2018; Langmead and Salzberg 2012). The tags with a mapping quality of at least 20 were selected for SNP calling. The SNPs called from the production step were filtered with three criteria: inbreeding coefficient of at least 0.8, Fisher Exact Test (p-value<0.001) to determine biallelic single locus SNPs (Poland et al., 2012) and Chi-square test for biallelic segregation with 96 percent expected inbreeding. SNPs that passed at least one filtering criteria were retained, and the subsequent SNP set was filtered to remove those with greater than 50 percent missing data and less than 0.01 MAF. Entries with over 50 percent missing data were also removed from the analysis. The final dataset

consisted of 13,271 SNP markers for 1,831 entries. Missing markers were imputed with the marker mean. The genomic relationship matrix (**G**) between entries was calculated according to Endelman and Jannink (2012). The additive pedigree matrix (**A**) was calculated as twice the coefficient of parentage.

Genetic Value Estimation

Genetic values of each trait for each entry were estimated for the YT and SP experiments separately within each breeding cycle, as well as for the SP_{BP} plots.

For the YT, best linear unbiased predictors (BLUPs) were calculated for the agronomic traits and for each VI at each time-point by fitting the following mixed model:

$$y_{ijkl} = \mu + g_i + t_j + r_{k(j)} + b_{l(jk)} + \varepsilon_{ijkl} \quad (2)$$

where y_{ijkl} is the trait value; μ is the overall mean; g_i is the random genetic effect for genotype i (BLUP), which are assumed to be independently and identically distributed according to a normal distribution $g_i \sim$iid $N(0, \sigma_g^2)$; $t_j \sim$iid $N(0, \sigma_t^2)$ is the random effect for trial j; $r_{k(j)} \sim$iid $N(0, \sigma_r^2)$ is the random effect for replicate k within trial j; $b_{l(jk)} \sim$iid $N(0, \sigma_b^2)$ is the random effect for block l within replicate k and trial j; and $\varepsilon_{ijkl} \sim$iid $N(0, \sigma_\varepsilon^2)$ is the residual effect. For DTHD and DTMT, which were evaluated on the first replicate only, the random effect for replicate was removed from model (2).

For the SP and SP_{BP} experiments, BLUPs were calculated for the agronomic traits and VIs as by fitting the model:

$$y_i = \mu + g_i + \varepsilon_i \quad (3)$$

where y_i is the trait value; μ is the overall mean; $g_i \sim$iid $N(0, \sigma_g^2)$ is the random genetic effect for genotype i; and $\varepsilon_i \sim$iid $N(0, \sigma_\varepsilon^2 \boldsymbol{R_i})$ is the residual variance for genotype i,

where $\boldsymbol{R_i}$ is the correlation matrix for residual effects parameterized as a one-dimensional autoregressive (AR1) process in the column direction to account for potential spatial correlation of observations. An AR1 process model was not applied in the direction of rows as full-sibs were sown adjacent to one another in rows, resulting in confounding between genetic and spatial variation in the row direction.

To account for the effect of phenology, BLUPs for GY and the VIs in the YT and SP were calculated a second time by adding a fixed effect covariate for DTHD to models (2) and (3). BLUPs for the VIs in the SP_{BP} were corrected for DTHD including the qualitative DTHD scores as a fixed effect covariate in model (3).

Significant outliers (p-value<0.05) were identified using Studentized Residuals and were removed from the analysis. In addition, the dataset was subset to remove two families that contained disproportionately high numbers of full-sibs. After further removing entries based on missing genotypic data, the final dataset included 839 and 920 entries in the 2016-17 and 2017-18 breeding cycles, respectively.

To avoid shrinkage on the same data twice (once during the calculation of iid BLUPs and again in the GS models), BLUPs were de-regressed by dividing by their reliability ($1 - \frac{\mathbf{PEV}}{\sigma_g^2}$), where **PEV** is the prediction error variance of the BLUP and σ_g^2 is the genotypic variance (Garrick *et al.,* 2009).

Weights for the error variances of the BLUPs were calculated as follows:

$$\frac{1}{\sqrt{\frac{1-H^2}{0.1+\frac{1-r^2}{r^2}H^2}}} \quad (4)$$

where H^2 is the proportion of the total variance explained by the genotypic variance component, σ_g^2, and r^2 are the reliabilities of the BLUPs (Garrick *et al.,* 2009). These

weights were used in downstream genomic prediction analyses to preserve information about heterogeneous variances driven by the differences in plot size and replication between the YT and SP. Without weights, this information would be otherwise ignored in a two-step genomic prediction procedure.

Heritability, Trait Correlation, and Response to Selection

To calculate the narrow-sense heritability, variance components for each of the agronomic and HTP traits were estimated according to the following model:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zg} + \mathbf{e} \quad (5)$$

where **y** is a vector of de-regressed BLUPs calculated from models (2) or (3); **X** is the fixed effects design matrix; **b** is a vector of fixed effects; **Z** is the random effects design matrix; **g** is a vector of random effects for genotype estimated assuming **g**~iid $N(0, \mathbf{G}\sigma_a^2)$ where **G** is the genomic relationship matrix and σ_a^2 is the additive genetic variance; and **e** is a vector of residuals where **e**~iid $N(0, \mathbf{I}\sigma_e^2)$, σ_e^2 is the residual variance, and **I** is the identity matrix. Narrow-sense heritability was calculated as $h^2 = \frac{\sigma_a^2}{\sigma_a^2+\sigma_e^2}$.

Trait phenotypic correlations were calculated as the Pearson's correlations between the iid BLUPs for each trait derived from models (2) and (3). Variance components for deriving the genetic correlations between each pair of traits were calculated according to the bivariate model:

$$\begin{bmatrix}\mathbf{y_1}\\ \mathbf{y_2}\end{bmatrix} = \begin{bmatrix}\mathbf{X} & \mathbf{0}\\ 0 & \mathbf{X}\end{bmatrix}\begin{bmatrix}\mathbf{u_1}\\ \mathbf{u_2}\end{bmatrix} + \begin{bmatrix}\mathbf{Z_1} & 0\\ 0 & \mathbf{Z_2}\end{bmatrix}\begin{bmatrix}\mathbf{a_1}\\ \mathbf{a_2}\end{bmatrix} + \begin{bmatrix}\boldsymbol{\varepsilon_1}\\ \boldsymbol{\varepsilon_2}\end{bmatrix} \quad (6)$$

where $\mathbf{y_1}$ and $\mathbf{y_2}$ are vectors of de-regressed BLUPs for each trait calculated from models (1) or (2); **X** is the fixed effects design matrix, which is the same for both

traits; $\mathbf{u_1}$ and $\mathbf{u_2}$ are vectors of fixed effects for each trait; $\mathbf{Z_1}$ and $\mathbf{Z_2}$ are the random effects design matrices for each trait; $\mathbf{a_1}$ and $\mathbf{a_2}$ are vectors of random effects for each trait; and $\boldsymbol{\varepsilon_1}$ and $\boldsymbol{\varepsilon_2}$ are vectors of residuals for each trait. The model was fit assuming $\begin{bmatrix}\mathbf{a_1}\\ \mathbf{a_2}\end{bmatrix} \sim N(0, \mathbf{G}\otimes\mathbf{H})$, where **G** is the genomic relationship matrix. For the residual, $\begin{bmatrix}\boldsymbol{\varepsilon_1}\\ \boldsymbol{\varepsilon_2}\end{bmatrix} \sim N(0, \mathbf{I}\otimes\mathbf{R})$, where **I** is the identity matrix and **R** is the residual variance-covariance matrix between the two traits. The resulting variance components were used to calculate the genetic correlation between each pair of traits according to:

$$r_{g_{12}} = \frac{\mathrm{cov}_{g_{12}}}{\sqrt{\mathrm{var}_{g_{11}}\mathrm{var}_{g_{22}}}} \quad (7)$$

where $r_{g_{12}}$ and $\mathrm{cov}_{g_{12}}$ are the genetic correlation and covariance, respectively, between the traits, and $\mathrm{var}_{g_{11}}$ and $\mathrm{var}_{g_{22}}$ are the genetic variances for each trait.

Response to selection was assessed by identifying the top performing individuals according to traits measured in the SP and calculating the difference in YT trait values between the mean of the those superior individuals and the total population. Response to selection was estimated for selection intensities from 100 (no selection) to 20 percent of individuals.

Selection Schemes

A train-test (TRN-TST) partitioning scheme was developed to estimate model accuracy within each breeding cycle. In each partition, a random 80 percent of the total number of individuals within a breeding cycle was sampled to form the TRN set, while the remaining 20 percent of entries formed the TST set. Sampling was performed on whole families to avoid full-sibs from the same family appearing in both the TRN and TST sets.

High-Throughput Phenotyping Selection: To assess the ability to predict GY of the YT using solely HTP information from the SP stage, the following linear regression model was fit using GY and HTP information from the YT for the entries in the TRN set:

$$y_i = \mu + z_i\alpha + \varepsilon_i \text{ (8)}$$

where y_i is the de-regressed BLUP of genotype i for GY in the YT calculated from model (2), μ is the overall mean, z_i is the de-regressed BLUP of genotype i for an HTP time-point observed in the YT calculated from model (2), α is the regression coefficient for the HTP trait, and ε_i is the residual for genotype i.

This model was then applied to predict GY for the individuals in the TRN set. The de-regressed BLUPs for the same HTP time-point observed in the SP calculated from model (3) were assigned to z_i while values for GY were "hidden." Models were assessed for each VI on each HTP time-point.

Univariate Genomic Selection: As a basis for comparison, univariate GS models using genomic marker or pedigree information were developed to predict GY for each breeding cycle following the form:

$$y_i = \mu + g_i + \varepsilon_i \text{ (9)}$$

where y_i are the de-regressed BLUPs for GY in the YT calculated in model (2); μ is the overall mean; g_i is the random effect for genotype i with $g_i \sim \text{iid } N(0, \mathbf{G}\sigma_g^2)$ or $g_i \sim \text{iid } N(0, \mathbf{A}\sigma_g^2)$, where **G** is the genomic relationship matrix and **A** is the pedigree matrix; and $\varepsilon_i \sim \text{iid } N(0, \mathbf{R}\sigma_\varepsilon^2)$ is the residual error for genotype i where **R** is the residual variance matrix. Weights from model (4) were included in the diagonal of **R**. These models contained no information on HTP traits.

Multi-Trait Genomic Selection: Finally, multi-trait GS approaches integrating HTP records with genomic/pedigree information to predict GY were developed following the bivariate model (6) where $\mathbf{y_1}$ is a vector of de-regressed BLUPs for GY in the YT calculated from model (2); $\mathbf{y_2}$ is a vector of de-regressed BLUPs for an HTP time-point observed in the YT calculated from model (2); **X** is the fixed effects design matrix, which is the same for each trait; $\mathbf{u_1}$ and $\mathbf{u_2}$ are vectors of fixed effects for GY and the HTP time-point; $\mathbf{Z_1}$ and $\mathbf{Z_2}$ are the random effects design matrices for GY and the HTP time-point; $\mathbf{a_1}$ and $\mathbf{a_2}$ are vectors of random effects for GY and the HTP time-point; and $\boldsymbol{\varepsilon_1}$ and $\boldsymbol{\varepsilon_2}$ are vectors of residuals for GY and the HTP trait. This model was fit assuming $\begin{bmatrix}\mathbf{a_1}\\ \mathbf{a_2}\end{bmatrix} \sim N(0, \mathbf{G}\otimes\mathbf{H})$ or $\begin{bmatrix}\mathbf{a_1}\\ \mathbf{a_2}\end{bmatrix} \sim N(0, \mathbf{A}\otimes\mathbf{H})$, where **G** is the genomic relationship matrix, **A** is the pedigree matrix, and **H** is the variance-covariance matrix for GY and the HTP time-point. The weights from model (4) were included in the diagonal of **R**.

This model was then applied to the individuals in the TST set, for which the de-regressed BLUPs for the same HTP time-point observed in the SP calculated from model (3) represented $\mathbf{y_2}$ while the values for GY were "hidden". All possible combinations of HTP time-points, VIs, and relationship matrices were tested.

The same 20 TRN-TST partitions were evaluated for HTP selection, univariate GS, and multi-trait GS to ensure fair comparisons between methods. Prediction models were validated by taking the Pearson's correlation of the predicted values for the TST set individuals with the iid BLUPs for GY in the YT calculated from model (2). Predictive abilities are represented by the mean and standard deviation of the correlations for the 20 TRN-TST partitions.

Accounting for Phenology

Due to the strong associations between GY, VIs, and phenology observed in this population, selection schemes based on the relationship between the VIs and GY would indirectly affect the distribution of phenology, disproportionally favoring later maturing entries. Although late maturity often confers superior yields, it is not always favorable as it may expose the crop to terminal and continual high temperature stress (Joshi et al., 2007; Mondal et al., 2013). In addition, breeders aim to maintain a range of maturity in the CIMMYT germplasm to allow the development of elite varieties for different growing regions. Therefore, two approaches were evaluated to account for the influence of phenology on GY.

Correcting Genetic Values for Phenology: In the first approach, iid BLUPs for DTHD observed in the YT calculated from model (2) were included in prediction models (8), (9), and (5) as a fixed effect covariate. For model validation, the predicted values were correlated to iid BLUPs for GY in the YT that had likewise been corrected for DTHD by including DTHD records observed in the YT as a fixed effect covariate in model (2). Prediction models (8), (9), and (5) were tested both with and without this correction for DTHD.

Selection Index for Grain Yield and Phenology: A selection index taking into account the genetic correlation between GY and DTHD was developed to assess the ability of prediction models (8), (9), and (5) to identify individuals with high GY without selecting directionally on DTHD. A restricted selection index aims to improve one or more traits while holding other traits constant (Kempthorn and Nordskog, 1959). The weights for a restricted selection index can be derived from a desired gains

index where the desired gain for the trait or traits to be held constant is set to zero (Openshaw and Hadley, 1984). The desired gains index takes the form:

$$\mathbf{b} = \mathbf{G^{-1}d} \text{ (10)}$$

where $\mathbf{b}$ is a vector of the selection index weights, $\mathbf{G^{-1}}$ is the inverse of the genetic variance-covariance matrix among traits, and $\mathbf{d}$ is a vector of the desired gains for the traits (Pešek and Baker, 1969).

To avoid directional selection on DTHD, the desired gain for DTHD, d_{DTHD}, was assigned to zero. In each breeding cycle, the desired gain for GY, d_{GY}, was set equal to one standard deviation of the iid BLUPs for GY in the YT. $\mathbf{G^{-1}}$ was calculated by fitting the bivariate model (5) with the genetic relationship matrix **G** and deregressed BLUPs for GY and DTHD in the YT. $\mathbf{G^{-1}}$ was calculated separately for each breeding cycle, using records for all individuals. The index weights, $\mathbf{b}$, were then derived in model (10) using the resulting values for $\mathbf{G^{-1}}$ and assigned values for $\mathbf{d}$.

To validate the prediction models in the context of the restricted selection index, models (8), (9), and (5) were fit as described to predict GY for the TST set. The GY predictions were then combined with observed records for DTHD in the SP to obtain the "predicted" restricted gains index values:

$$I_i = b_{GY}X_{GY_i} + b_{DTHD}X_{DTHD_i} \text{ (11)}$$

where I_i is the index value for genotype *i*; b_{GY} and b_{DTHD} are the index weights for GY and DTHD, respectively, derived from model (10); X_{GY_i} is the predicted value for GY of genotype *i* from prediction models (8), (9), and (5); and X_{DTHD_i} is the observed iid BLUP for DTHD in the SP for genotype *i* (Smith, 1936; Hazel, 1943).

For model validation, the observed index values for all individuals were calculated with model (11) using the observed iid BLUPs for GY and DTHD in the YT derived from model (2). Predictive abilities are represented as the Pearson's correlations between the predicted and observed index values for the individuals in the TST set.

Software

Image processing was primarily conducted using Python, open source software QGIS (QGIS Development Team, 2019, QGIS Geographic Information System, Open Source Geospatial Foundation Project, http://qgis.osgeo.org), and Agisoft PhotoScan (Agisoft LLC, St. Petersburg, Russia). Data analyses were implemented in the R environment (R Core Team, 2016, Vienna, Austria) with the package "ASReml-R" (Gilmour et al., 2014) for R. The genomic relationship matrix (**G**) was calculated using the A.mat() function within the R package "rrBLUP" (Endelman 2011). The coefficients of parentage for calculating the pedigree relationship matrix (**A**) were estimated with the "Browse" application of the International Crop Information System software (McLaren et al., 2000).

Results

To evaluate the potential of utilizing aerial HTP to improve selection at the seed-limited stage of a bread wheat breeding program, we deployed a UAV equipped with a multi-spectral camera to phenotype vegetation indices for a set of 2,016 individuals sown both as unreplicated 1m × 0.8m SP and as twice-replicated 2.8m × 1.6m double bed YT during the 2016-17 and 2017-18 breeding cycles at CIMMYT in Ciudad Obregón, Sonora, México.

The distributions of BLUPs for GY in the YT were consistent across the two breeding cycles. The means and standard deviations for GY in the YT were 6.44±0.33 t ha^{-1} in 2016-17 and 6.24±0.32 t ha^{-1} in 2017-18 (Table 2.1). Values for GY in the SP were more variable within and across cycles. The means ± the standard deviations for GY in the SP were 5.20±0.64 t ha^{-1} in 2016-17 and 7.69±0.45 t ha^{-1} in 2017-18. Both the YT and SP reached the heading and maturity stages around the same time during both breeding cycles (Figure 2.1).

A UAV equipped with a multispectral camera was deployed to phenotype NDVI, GNDVI, and RENDVI during the grain filling stage between heading and maturity (Figure 2.1). The YT and SP were phenotyped at two time-points (TP-1: 13 Mar; TP-2: 20 Mar) in 2016-17 and at three time-points (TP-1: 07 Mar; TP-2: 12 Mar in the SP, 13 Mar in the YT; TP-3: 19 Mar) in 2017-18. The SP_{BP} was phenotyped with the UAV on the same time-points as the 2016-17 YT and SP.

Trait Heritabilities

Heritability estimates for GY were greater by 0.18 and 0.35 for the YT than for the SP in 2016-17 and 2017-18, respectively (Table 2.2). Correcting for DTHD reduced heritabilities for GY in the YT by 0.09 and 0.07 in 2016-17 and 2017-18, respectively. In the SP, correcting for DTHD reduced the heritability for GY by 0.06 in 2017-18 but had no effect in 2016-17. For a given VI time-point, the heritability estimates for the three VIs differed from one another by an average of 0.03. The greatest difference between VIs collected for a single time-point was observed between GNDVI ($h^2 = 0.85$) and RENDVI ($h^2 = 0.76$) in the YT on 07 Mar 2018 for a difference of 0.09. Due to the similarity between VIs for heritability, only the

Table 2.1: Trait correlations within and between the YT, SP, and SP_{BP}

2016-17		YT					SP				
		GRYLD	DTHD	DTMT	NDVI TP-1	NDVI TP-2	GRYLD	DTHD	DTMT	NDVI TP-1	NDVI TP-2
YT	GRYLD	6.45 (0.32)	0.41 (0.61)	0.44 (0.62)	0.50 (0.72)	0.51 (0.72)	0.32 (0.67)	0.40 (0.63)	0.42 (0.69)	0.43 (0.72)	0.44 (0.73)
	DTHD	-	74.5 (4.2)	0.78 (0.93)	0.74 (0.82)	0.80 (0.89)	0.08 (0.12)	0.82 (0.96)	0.65 (0.85)	0.60 (0.62)	0.71 (0.81)
	DTMT	-	-	122.5 (3.1)	0.72 (0.80)	0.80 (0.90)	0.12 (0.17)	0.73 (0.89)	0.69 (0.91)	0.59 (0.67)	0.72 (0.86)
	NDVI TP-1	0.32 (0.48)	-	-	0.824 (0.021)	0.94 (0.95)	0.17 (0.19)	0.72 (0.75)	0.69 (0.79)	0.75 (0.89)	0.78 (0.86)
	NDVI TP-2	0.34 (0.45)	-	-	0.85 (0.79)	0.669 (0.053)	0.14 (0.05)	0.78 (0.82)	0.76 (0.86)	0.76 (0.88)	0.84 (0.92)
SP	GRYLD	0.30 (0.78)	-	-	0.14 (0.41)	0.09 (0.21)	5.20 (0.64)	0.12 (0.05)	0.14 (0.12)	0.25 (0.27)	0.20 (0.16)
	DTHD	-	-	-	-	-	-	78.2 (3.5)	0.76 (0.85)	0.70 (0.64)	0.81 (0.81)
	DTMT	-	-	-	-	-	-	-	121.3 (3.5)	0.75 (0.84)	0.86 (0.95)
	NDVI TP-1	0.22 (0.54)	-	-	0.48 (0.89)	0.44 (0.91)	0.23 (0.33)	-	-	0.774 (0.038)	0.91 (0.93)
	NDVI TP-2	0.17 (0.47)	-	-	0.42 (0.74)	0.49 (0.89)	0.17 (0.21)	-	-	0.82 (0.90)	0.569 (0.082)

Table 2.1 (Continued)

2017-18		YT						SP						SP_{BP}	
		GRYLD	DTHD	DTMT	NDVI TP-1	NDVI TP-2	NDVI TP-3	GRYLD	DTHD	DTMT	NDVI TP-1	NDVI TP-2	NDVI TP-3	NDVI TP-1	NDVI TP-2
YT	GRYLD	6.24 (0.32)	0.43 (0.60)	0.45 (0.64)	0.46 (0.64)	0.45 (0.60)	0.49 (0.63)	0.40 (0.89)	0.42 (0.57)	0.39 (0.58)	0.37 (0.55)	0.39 (0.60)	0.42 (0.57)	0.33 (0.53)	0.36 (0.52)
	DTHD	-	79.0 (3.6)	0.84 (0.90)	0.73 (0.86)	0.79 (0.88)	0.80 (0.89)	0.19 (0.52)	0.88 (0.96)	0.66 (0.85)	0.67 (0.76)	0.73 (0.83)	0.80 (0.88)	0.51 (0.65)	0.61 (0.73)
	DTMT	-	-	126.1 (2.6)	0.69 (0.83)	0.74 (0.85)	0.79 (0.89)	0.16 (0.46)	0.76 (0.85)	0.68 (0.89)	0.63 (0.71)	0.66 (0.79)	0.74 (0.84)	0.51 (0.67)	0.63 (0.77)
	NDVI TP-1	0.23 (0.31)	-	-	0.491 (0.022)	0.88 (0.95)	0.87 (0.93)	0.25 (0.65)	0.67 (0.82)	0.57 (0.81)	0.72 (0.90)	0.73 (0.92)	0.71 (0.88)	0.51 (0.72)	0.56 (0.73)
	NDVI TP-2	0.19 (0.21)	-	-	0.72 (0.83)	0.504 (0.023)	0.89 (0.94)	0.23 (0.53)	0.72 (0.85)	0.62 (0.83)	0.73 (0.87)	0.77 (0.92)	0.78 (0.92)	0.62 (0.78)	0.65 (0.79)
	NDVI TP-3	0.24 (0.28)	-	-	0.69 (0.70)	0.71 (0.74)	0.355 (0.035)	0.24 (0.66)	0.75 (0.86)	0.67 (0.89)	0.69 (0.81)	0.74 (0.88)	0.79 (0.92)	0.56 (0.73)	0.64 (0.81)
SP	GRYLD	0.35 (0.84)	-	-	0.15 (0.43)	0.13 (0.22)	0.13 (0.52)	7.69 (0.45)	0.17 (0.60)	0.17 (0.57)	0.20 (0.58)	0.23 (0.60)	0.23 (0.88)	0.22 (0.46)	0.23 (0.44)
	DTHD	-	-	-	-	-	-	-	77.1 (3.8)	0.73 (0.91)	0.70 (0.79)	0.76 (0.84)	0.85 (0.60)	0.49 (0.67)	0.60 (0.74)
	DTMT	-	-	-	-	-	-	-	-	125.8 (2.4)	0.65 (0.83)	0.71 (0.89)	0.81 (0.84)	0.53 (0.80)	0.63 (0.88)
	NDVI TP-1	0.09 (0.24)	-	-	0.46 (0.80)	0.46 (0.73)	0.32 (0.51)	0.11 (0.26)	-	-	0.534 (0.021)	0.89 (0.95)	0.82 (0.89)	0.53 (0.72)	0.56 (0.70)
	NDVI TP-2	0.09 (0.28)	-	-	0.42 (0.78)	0.48 (0.82)	0.34 (0.61)	0.15 (0.26)	-	-	0.78 (0.90)	0.491 (0.027)	0.89 (0.95)	0.58 (0.78)	0.62 (0.79)
	NDVI TP-3	0.09 (0.19)	-	-	0.28 (0.60)	0.42 (0.76)	0.40 (0.76)	0.15 (0.18)	-	-	0.59 (0.61)	0.73 (0.82)	0.397 (0.047)	0.62 (0.73)	0.71 (0.86)
SP_{BP}	NDVI TP-1	0.11 (0.26)	-	-	0.17 (0.33)	0.34 (0.50)	0.20 (0.31)	0.12 (0.27)	-	-	0.21 (0.34)	0.28 (0.52)	0.34 (0.54)	0.759 (0.035)	0.85 (0.95)
	NDVI TP-2	0.09 (0.20)	-	-	0.13 (0.22)	0.27 (0.39)	0.22 (0.37)	0.11 (0.21)	-	-	0.14 (0.26)	0.21 (0.43)	0.34 (0.62)	0.77 (0.90)	0.549 (0.068)

Table 2.1 (Continued)

YT, yield trials; SP, small plots; SP_{BP}, small plots from the 2016-17 formal breeding program; GY, grain yield; DTHD, days to heading; DTMT, days to maturity; NDVI, normalized difference vegetation index; TP, time-point. The diagonals contain the mean of the iid BLUPs for each trait derived from models (1) and (2). The values in parentheses along the diagonal are the standard deviations of the iid BLUPs for each trait. The upper triangles contain the Pearson's correlation between each pair of traits. The lower triangles contain the Pearson's correlations among and between GY and the NDVI time-points following correction for DTHD. The values in parentheses in the upper and lower triangles are the genetic correlations between each pair of traits.

Figure 2.1: UAV phenotyping time-points in relation to heading and maturity dates in each experiment

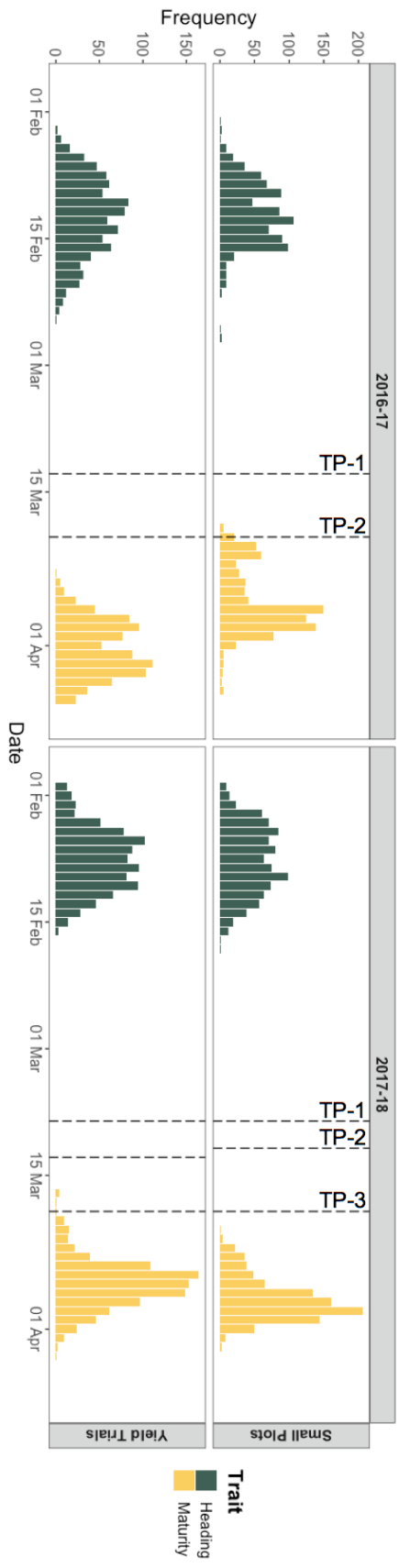

Frequency
0
50
100
150
200
2016-17
2017-18
TP-1
TP-2
TP-3
01 Feb
15 Feb
01 Mar
15 Mar
01 Apr
Date
Small Plots
Yield Trials
Trait
Heading
Maturity

Figure 2.1 (Continued)

YT, yield trials; SP, small plots; TP, time-point; UAV, unmanned aerial vehicle; VI, vegetation index. The histograms represent the heading and maturity dates of the SP and YT in 2016-17 and 2017-18. The dotted red lines indicate the time-points on which each experiment was phenotyped for VIs with the UAV.

Table 2.2: Narrow-sense heritability (h^2) estimates for the agronomic and VI traits

Breeding Cycle	Trait	YT h^2	SP h^2	SP_{BP} h^2
2016-17	GY	0.55 (0.46)	0.39 (0.40)	-
	DTHD	0.79	0.82	-
	DTMT	0.65	0.80	-
	NDVI TP-1	0.77 (0.58)	0.69 (0.75)	0.79 (0.65)
	NDVI TP-2	0.84 (0.57)	0.79 (0.78)	0.83 (0.65)
2017-18	GY	0.72 (0.65)	0.37 (0.31)	-
	DTHD	0.82	0.85	-
	DTMT	0.83	0.65	-
	NDVI TP-1	0.79 (0.59)	0.82 (0.74)	-
	NDVI TP-2	0.79 (0.58)	0.78 (0.65)	-
	NDVI TP-3	0.83 (0.64)	0.83 (0.59)	-

YT, yield trials; SP, small plots; SP_{BP}, small plots from the 2016-17 formal breeding program; GY, grain yield; DTHD, days to heading; DTMT, days to maturity; NDVI, normalized difference vegetation index; TP, time-point. Values in parentheses are the narrow-sense heritabilities for the traits corrected for DTHD.

estimates for NDVI are shown in Table 2.1.

Trait Phenotypic and Genetic Correlations

Strong phenotypic and genetic correlations were observed among and between the agronomic and VI traits within the YT, SP, and SP_{BP} (Table 2.1). The phenotypic and genetic correlations for each pair of traits largely followed the same trends, and therefore only the phenotypic correlations will be discussed herein. The highest correlations were observed between the three VIs recorded on the same time-point. The average correlation between VIs taken on the same time-point was 0.90. Due to the high correlation between VIs, correlations are shown in Table 2.1 for NDVI only. High correlations were also observed for NDVI recorded on different time-points. These ranged from 0.82 to 0.94 with an average of 0.89. The correlation between GY measured in the YT in SP was 0.32 in 2016-17 and 0.40 in 2017-18.

In the YT, the correlations between the NDVI time-points and GY ranged from 0.45 to 0.51, while in the SP, correlations were slightly lower with a range of 0.20 to 0.25. Correlations from 0.37 to 0.44 were observed between NDVI time-points evaluated in the SP and GY measured in the YT. For NDVI measured in the SP_{BP}, correlations with GY in the YT were 0.33 for TP-1 and 0.36 for TP-2.

Correcting for DTHD had a variable impact on the correlation between NDVI and GY. In the YT, correcting for DTHD reduced the correlations between the NDVI time-points and GY by an average of 0.22. The correction had a smaller effect on the relationship between NDVI and GY in the SP, where it reduced correlations by 0.06. When NDVI and GY were observed in the SP and YT, respectively, correcting for DTHD reduced the correlations by an average of 0.28.

Strong relationships were also observed between the NDVI time-points and DTHD/DTMT, averaging 0.76 for the YT and SP. Notably, DTHD/DTMT observed in both the YT and SP was moderately predictive of GY in the YT, with correlations ranging from 0.39 to 0.45.

Selection on High-Throughput Phenotyping Traits

Potential selection scenarios utilizing SP or SP_{BP} traits as the selection criteria to improve GY in the YT are illustrated in Figure 2.2. On average, selecting on NDVI at the SP stage would have increased GY in the YT by 2.17 percent, 1.68 percent, and 0.97 percent at intensities of 25 percent, 50 percent, and 75 percent, respectively. By comparison, selecting on GY at the SP stage would have increased GY in the YT by 2.20 percent, 1.56 percent, and 0.83 percent, on average, at the respective intensities. While selecting on GY at the SP stage would have had little effect on DTHD in the YT, selection on NDVI would have favored later maturing individuals, particularly at high selection intensities. The average DTHD in the YT would have increased by 3.2, 2.2, and 1.2 days at selection intensities of 25 percent, 50 percent, and 75 percent.

When correcting the NDVI records in the SP for DTHD, the response to selection for GY at the YT stage was reduced to below 1 percent at most levels of selection intensity. Correcting the GY records in the SP for DTHD also reduced the projected response to selection for GY in the YT, however the decrease was less pronounced. As expected, selecting on DTHD-corrected NDVI or GY records from the SP would have had minimal impact on the distribution of DTHD in the YT.
By selecting on NDVI time-points recorded for the SP_{BP}, GY in the YT would have increased by 2.01 percent, 1.31 percent, and 0.83 percent on average at selection

Figure 2.2: Response to selection in the YT when selecting on traits measured in the SP and SP_{BP}

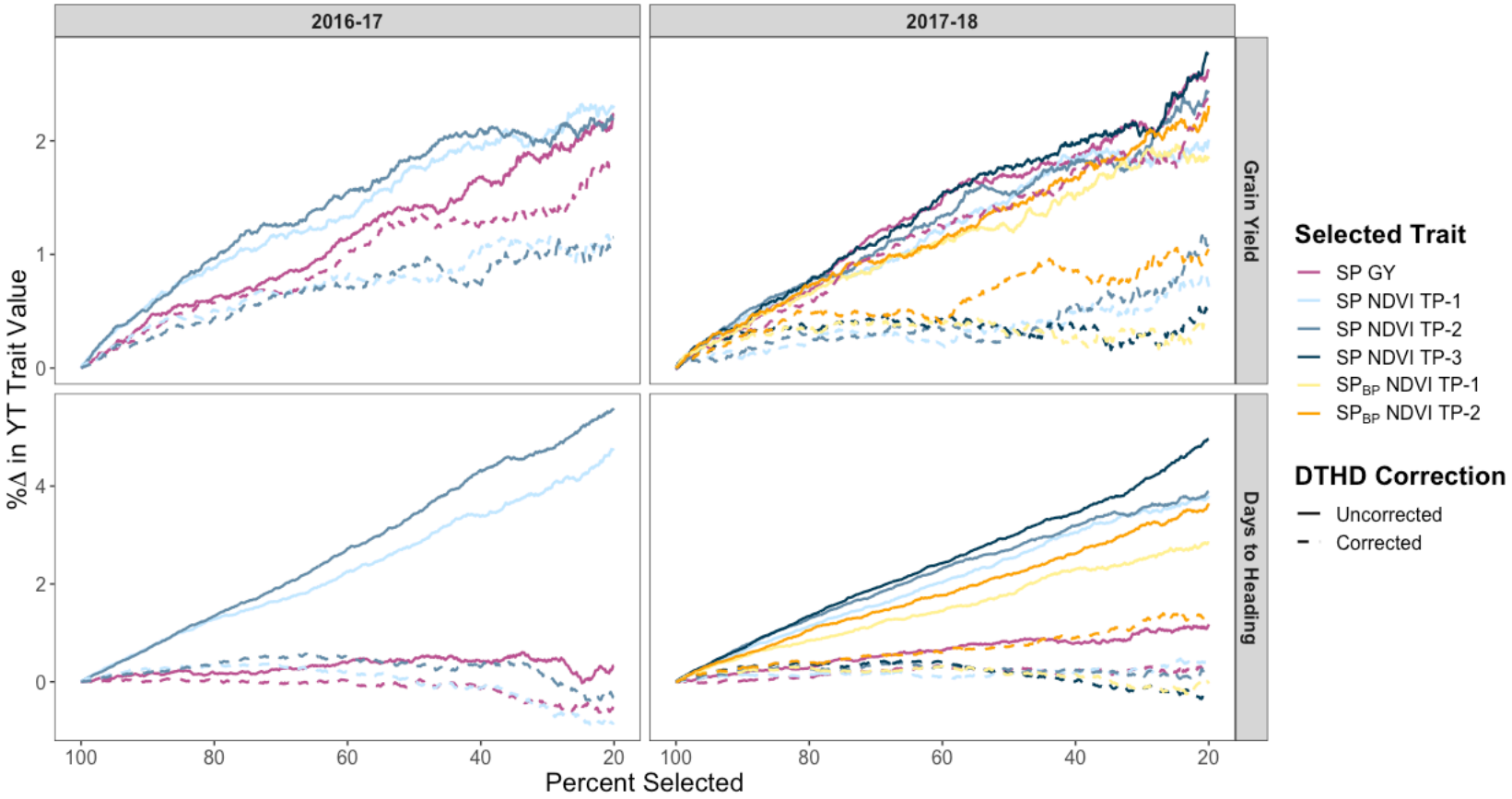

2016-17
2017-18
Grain Yield
Days to Heading
%Δ in YT Trait Value
Percent Selected
100
80
60
40
20
0
1
2
4
Selected Trait
SP GY
SP NDVI TP-1
SP NDVI TP-2
SP NDVI TP-3
SP$_{BP}$ NDVI TP-1
SP$_{BP}$ NDVI TP-2
DTHD Correction
Uncorrected
Corrected

Figure 2.2 (Continued)

SP_{BP}, small plots from the 2016-17 formal breeding program; GY, grain yield; NDVI, normalized difference vegetation index; DTHD, days to heading. The x-axis represents the selection intensity expressed as the percentage of the population selected based on the SP or SP_{BP} trait. The y-axis represents the corresponding percent change in the YT trait value.

intensities of 25 percent, 50 percent, and 75 percent. As in the SP, high selection intensity on NDVI time-points in the SP_{BP} would have led to an increase in DTHD in the YT.

Visual Versus High-Throughput Phenotyping Selection

Prior to harvest, the breeders made visual selections within the first 300 plots of the SP_{BP} for promotion to the first-year YT in the following breeding cycle. Of the 293 individuals (excluding checks), 26 were selected, representing a selection intensity of 12.3 percent. Visual selection resulted in a 0.38 percent response to selection in GY of the SP_{BP}.

Among the total population of 293 individuals, 187 (63.8 percent) were scored as "early," 51 (17.4 percent) as "mid," 49 (16.7 percent) as "late," and five (1.7 percent) as "very late" for DTHD. For the 36 individuals that were visually selected by the breeders, 22 (61.1 percent) were scored as "early," eight (22.2 percent) as "mid," and six (16.7 percent) as "late" for DTHD. Although the breeders eliminated the "very late" material through visual selection, the proportions of selected individuals belonging to the other DTHD subgroups were comparable to those of the total population of 293 individuals.

In the SP_{BP}, the NDVI time-points had correlations with GY of 0.13 for TP-1 and 0.05 for TP-2 (Figure 2.3A). Despite these weak relationships, NDVI was effective in identifying SP_{BP} individuals with low GY. If the same level of selection intensity had been applied to the SP_{BP} using the NDVI time-points as the selection criteria, the response to selection in GY of the SP_{BP} would have been 1.54 percent and 1.40 percent for TP-1 and TP-2, respectively. Of the top 36 individuals in terms of

Figure 2.3: Visual versus HTP selection in the SP_{BP}

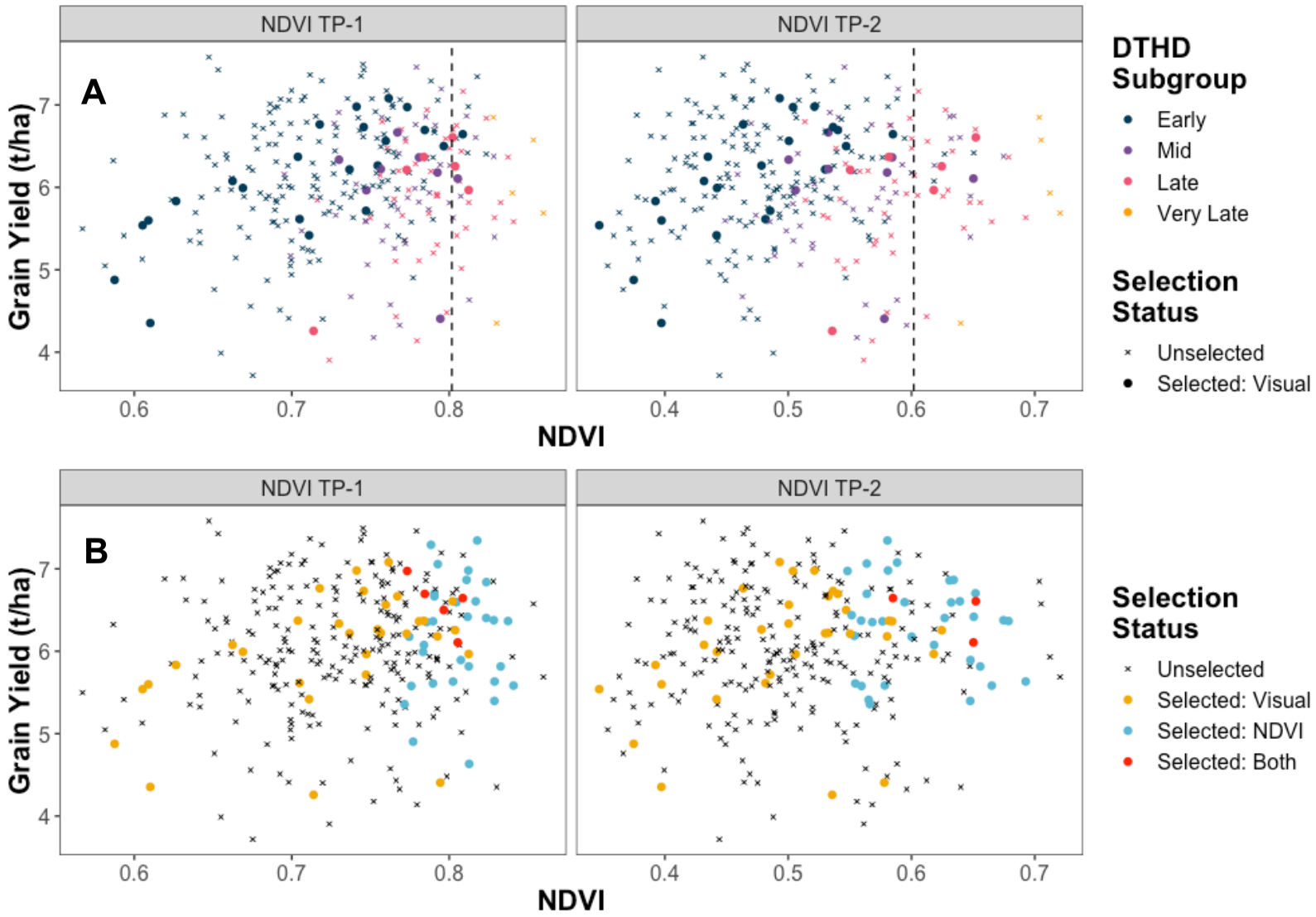

NDVI TP-1
NDVI TP-2
A
B
Grain Yield (t/ha)
NDVI
DTHD Subgroup
Early
Mid
Late
Very Late
Selection Status
Unselected
Selected: Visual
Selected: NDVI
Selected: Both

Figure 2.3 (Continued)

The x-axis shows NDVI values measured in the SP_{BP}. The y-axis represents GY values assessed in the SP_{BP}.

A. The color represents the qualitative DTHD subgroups the individuals were assigned to. The shape indicates which individuals were visually selected by breeders. Points to the right of the dashed lines indicate individuals that would have been selected using NDVI as the selection criteria, given the same level of selection intensity used by the breeders for visual selection.

B. Directional selection on DTHD is avoided by using NDVI to select the top 22, 8, and 6 individuals in the “early”, “mid”, and “late” DTHD subgroups, respectively. The color and shape indicate if which individuals were selected, and, if so, by which selection method(s).

NDVI, 25 would have been selected by both TP-1 and TP-2. Only five and four of the individuals visually selected by breeders were among the top 36 in terms of NDVI TP-1 and TP-2, respectively. Among the top 36 individuals for NDVI TP-1, six (16.7 percent) were scored as "early", eight (22.2 percent) as "mid", 17 (47.2 percent) as "late", and five (13.9 percent) as "very late" for DTHD. For NDVI TP-2, three (8.3 percent) of the top 36 individuals were scored as "early", nine (25.0 percent) were scored as "mid", 19 (52.8 percent) were scored as "late", and five (13.9 percent) were scored as "very late", indicating considerable directional selection on DTHD.

Since CIMMYT breeders attempt to avoid selecting directionally on DTHD when performing visual selection, a more appropriate comparison might be to identify the top individuals in terms of NDVI for each qualitative DTHD score. The breeders visually selected 22, eight, and six individuals scored as "early," "mid," and "late," respectively. By taking subsets of the "early," "mid," and "late" individuals and identifying the top 22, eight, and six individuals, respectively, based on NDVI, the response to selection in GY for the SP_{BP} would have been 2.91 percent for TP-1 and 4.23 percent for TP-2 (Figure 2.3B). Using this approach, of the 36 individuals selected by NDVI, 24 would have been selected by both TP-1 and TP-2. Only five and three of the individuals visually selected by breeders were among the 36 selected by NDVI TP-1 and TP-2, respectively.

Predictive Abilities

Predictive abilities were assessed with a TRN-TST scheme that utilized YT records from 80 percent of entries as the TRN set to build prediction models. To avoid inflated accuracies, whole families were sampled at random such that there were full-

sibs from the same family common to both the TRN and TST sets. Predictions were then generated using the trained model for the remaining 20 percent of entries from the TST set. For HTP selection and multi-trait GS, which utilized HTP traits, the HTP records from the YT for the 80 percent of individuals in the TRN set were used to train the model, while the HTP records from the SP for the remaining 20 percent of individuals in the TST were used to predict. Two approaches were taken to address confounding between GY, DTHD, and NDVI.

Correcting Genetic Values for Phenology: In the first approach, iid BLUPs for DTHD in the YT were included as a fixed effect in the prediction model. Figure 2.4 shows the Pearson's correlations between the predictions and iid BLUPs for GY in the YT with and without correction for DTHD.

The predictive abilities of the three VIs gave similar results with no individual VI providing a consistent, significant advantage over the other VIs. On average, the difference in accuracy between the same models differing in only the VI used was 0.04. For this reason, only the predictive abilities of NDVI are shown in Figure 2.4.

In addition, the predictive abilities of models using the pedigree relationship matrix were low, with an average of 0.16 for the univariate GS models, likely due to the large numbers of full-sibs and missing information in the pedigree records. Therefore, the predictive abilities of univariate and multi-trait GS models shown in Figure 2.4 and discussed hereafter refer to those estimated using the genomic relationship matrix (**G**) only.

When using NDVI traits alone to predict yield in the absence of genomic marker or pedigree information, predictive abilities averaged 0.43 in 2016-17 and 0.38

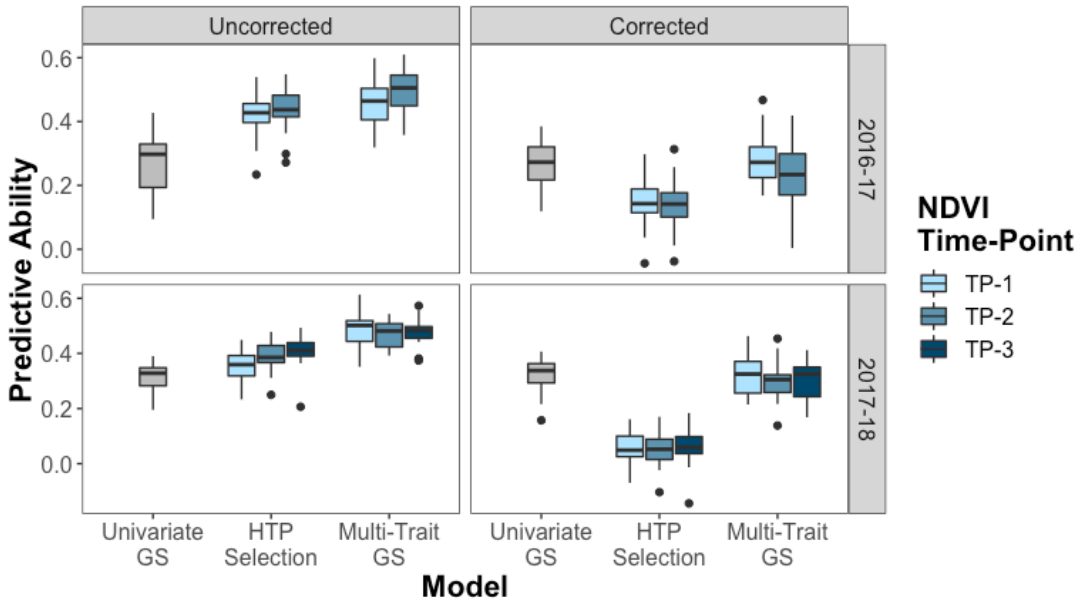


Figure 2.4: Predictive abilities of univariate GS, HTP selection, and multi-trait GS uncorrected and corrected for phenology

GS, genomic selection; HTP, high-throughput phenotyping. Predictive ability is expressed as the Pearson's correlations between predictions and observed iid BLUPs for GY in the YT across 20 random TRN-TST iterations. The univariate and multi-trait GS models were fit using the genomic relationship matrix (**G**). Corrected predictive abilities were estimated using genetic values that were calculated with DTHD as a fixed effect covariate.

in 2017-18 with little variation observed between NDVI time-points. Correcting for DTHD reduced the predictive abilities to averages of 0.14 and 0.05 for the 2016-17 and 2017-18 breeding cycles, respectively.

In the absence of NDVI phenotypes, the predictive abilities for the univariate GS models were 0.27 and 0.31 for the 2016-17 and 2017-18 breeding cycles, respectively. Before correcting for DTHD, the HTP selection models exceeded the univariate GS models in terms of predictive ability averages of 0.16 and 0.07 in 2016-17 and 2017-18, respectively. However, since the correction for DTHD had a strong effect on predictive abilities of the HTP selection models but negligible impact on univariate GS, the predictive ability of univariate GS exceeded that of the NDVI-based models by averages of 0.13 and 0.27 in 2016-17 and 2017-18, respectively, following the correction for DTHD.

The multi-trait GS models that combined genomic marker information with HTP records were observed to have an average predictive ability of 0.48. In this case, the multi-trait GS models gave the highest accuracies of all, exceeding the HTP selection models by averages of 0.05 and 0.09 and the univariate GS models by 0.21 and 0.17 for breeding cycles 2016-17 and 2017-18, respectively. However, following the correction for DTHD, the predictive abilities of the multi-trait GS models were more similar to those of univariate GS.

Selection Index for Grain Yield and Phenology: To evaluate the ability to use NDVI, both independently and in combination with genetic information, to identify superior individuals for GY without selecting directionally on DTHD, a restricted

selection index was developed to take into account the genetic correlation between GY and DTHD. The desired gain for DTHD, d_{DTHD}, was set to 0 for both breeding cycles to restrict selection on DTHD. The desired gain for GY, d_{GY}, was set to one standard deviation of the iid BLUPs for GY in the YT, which amounted to 0.325 and 0.316 for 2016-17 and 2017-18, respectively. A bivariate model for GY and DTHD in the YT was fit with the genomic relationship matrix **G**, and the resulting variance-covariance matrix was used to calculate the selection index weights, which amounted to 6.308 and 3.927 for b_{GY} and -0.224 and -0.183 for b_{DTHD} in 2016-17 and 2017-18, respectively.

Observed index values for all individuals were calculated using the index weights and the observed iid BLUPs for GY and DTHD in the YT. The top 10 percent of individuals in terms of the observed index are shown with respect to GY and DTHD in Figure 2.5. The observed index values showed phenotypic correlations of 0.89 and 0.85 with GY in 2016-17 and 2017-18, respectively, conferring a response to selection of 0.44 in both breeding cycles. The index had a negligible relationship with DTHD, with correlations of -0.05 and -0.10 and responses to selection of -1.2 and -0.6 in 2016-17 and 2017-18, respectively. The narrow-sense heritabilities of the index were 0.47 in 2016-17 and 0.64 in 2017-18. Genetic correlations between the index and the NDVI traits in the SP were 0.47 and 0.38 for TP-1 and TP-2, respectively, in 2016-17 and 0.13, 0.14, and 0.05 for TP-1, TP-2, and TP-3, respectively, in 2017-18.

Predictions for GY of the TST set from the HTP selection, univariate GS, and multi-trait GS were used with observed iid BLUPs for DTHD in the SP and index weights to develop the index predictions. The predictive abilities of each model in terms of the index are shown in Figure 2.6. The predictive abilities for HTP selection

Figure 2.5: Restricted selection index to increase GY while holding DTHD constant

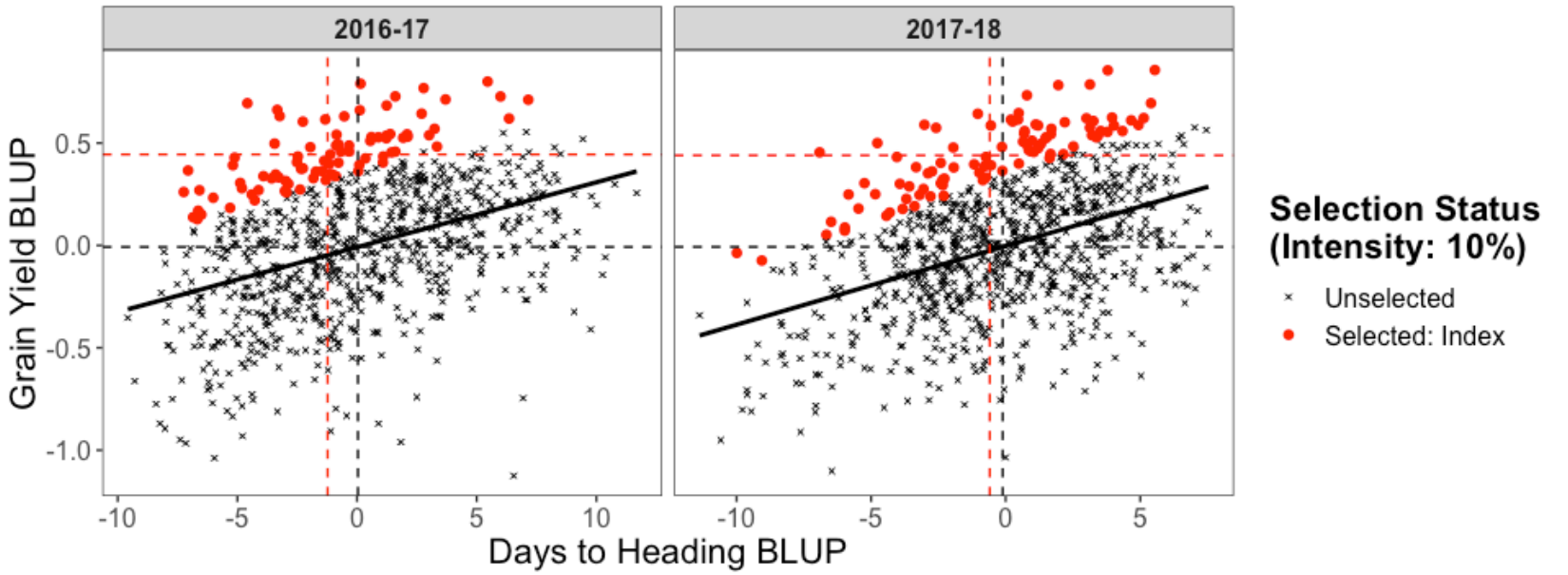

2016-17
2017-18
Grain Yield BLUP
0.5
0.0
-0.5
-1.0
-10
-5
0
5
10
Days to Heading BLUP
Selection Status
(Intensity: 10%)
Unselected
Selected: Index

Figure 2.5 (Continued)

Plotted points are the iid BLUPs for DTHD and GY in the YT in 2016-17 and 2017-18. The color and shape of the points indicate which individuals were selected based on the restricted selection index at an intensity of 10 percent. The solid black line is the regression line. The dashed black lines indicate the trait means for GY and DTHD. The dashed red lines indicate the response to selection in both traits.

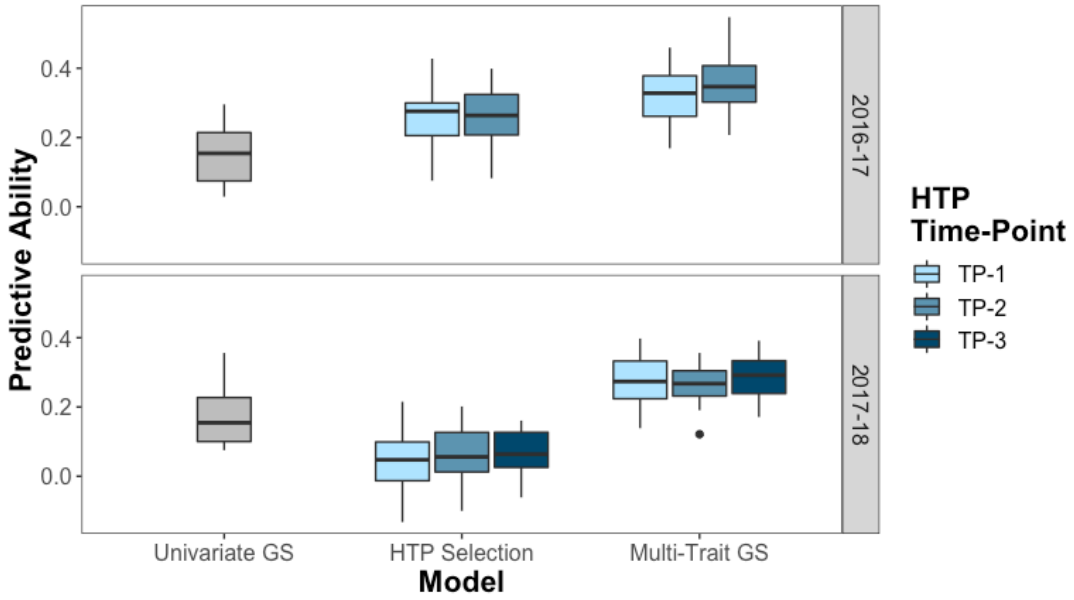


Figure 2.6: Predictive abilities of univariate GS, HTP selection, and multi-trait GS for a restricted selection index to increase GY while holding DTHD constant

Predicted index values for the TST set were calculated using the predicted values for GY and observed iid BLUPs for DTHD in the SP. The predictive ability is expressed as the Pearson's correlations between predicted values and observed index values calculated using the observed iid BLUPs for GY and DTHD in the YT across 20 random TRN-TST iterations. The univariate and multi-trait GS models were fit using the genomic relationship matrix (**G**).

and univariate GS were similar in 2016-17; however, in 2017-18 the predictive ability of univariate GS exceeded HTP selection by 0.14. Multi-trait GS gave the highest predictive abilities during both breeding cycles. In 2016-17, the multi-trait GS predictive abilities exceeded those of HTP selection and univariate GS by an average of 0.11. In 2017-18, average increases of 0.20 and 0.07 over HTP selection and univariate GS, respectively, were observed.

Discussion

The aerial HTP traits were found to be more heritable than GY in the YT and SP, and minimal differences were observed with respect to heritability among the multiple time-points on which the HTP traits were recorded. Several previous studies have found HTP traits to be heritable and repeatable (Rutkoski et al., 2014; Crain et al., 2017; Sun et al., 2017), though the traits measured in these studies were evaluated on replicated trials with plot sizes sufficient for reliably assessing yield. To our knowledge, this is the first report aiming to assess the heritability of HTP traits measured on small, unreplicated plots and estimate the extent to which they are predictive of GY in replicated yield trials.

It is possible, however, that the estimates of narrow-sense heritability for traits measured in the SP are partially inflated. At the early stages of small grains breeding programs, breeders often sow full-sibs derived from the same cross adjacent to one another in the field for logistical ease and to enable visual comparisons among individuals within families to be made. The disadvantage of this approach, however, is that the genetic and spatial components of the phenotypic variance are confounded.

When planting siblings together in families, Magnussen (1993) showed that the estimates of additive genetic variance and narrow-sense heritability were biased upward with positive spatial autocorrelation among neighboring plots. In this study, the small plot size and lack of replication in the SP may have further contributed to the imprecision of variance component estimation. Without randomized field designs, it is difficult to achieve reliable estimates of heritability and genetic correlations, particularly when genotypes is unreplicated.

Despite this, the phenotypic and genetic correlations between the HTP traits measured in the SP and GY in the YT were relatively high. The second replicate of the YT was randomized, and the YT and SP also represented separate field experiments, each with independent environmental effects. In addition, the HTP traits in the SP_{BP} were also correlated with GY in the YT, despite having been measured during the prior breeding cycle and on smaller plot sizes than the SP. Even if heritability estimates measured at the unreplicated early generation stages are imprecise, it appears that selection on these traits may still be effective for improving GY at the YT stage.

The current selection method used at the early generation seed-limited stage of the CIMMYT irrigated bread wheat breeding program is to first apply visual selection on the plots based on agronomic characters such as plant height, phenology, disease resistance, vigor, etc., and then to perform a second round of culling after harvest based on visual assessments of the seed grain quality (van Ginkel et al., 2002). Harvesting 300 plots of the SP_{BP} for GY enabled a direct comparison between HTP and the visual selection scheme because all individuals, both selected and unselected

through visual selection, were assessed for both GY and HTP. Although the correlations between HTP and GY in the SP_{BP} were weak, the HTP traits were successful in identifying individuals with low GY. Using HTP as the selection criterion was therefore more effective than visual selection in terms of response to selection for GY.

While these results are encouraging, the sample size of this experiment was relatively small. In addition, the GY values were based on the harvest of unreplicated 0.8m × 0.8m plots, and the YT and SP experiments demonstrated that GY recorded in the SP is only moderately predictive of GY at the YT stage. It is difficult to conduct research on response to selection for different selection methods within active breeding programs because breeders do not evaluate unselected material in replicated yield trials as is needed for assessment of the unselected population mean. For example, in this study, the breeders had already visually selected the individuals tested in the YT and SP experiments out of the SP_{BP} during the breeding cycle prior. Therefore, any SP_{BP} entries that were not selected were not evaluated in the YT during the following cycle, and the unselected population mean could not be estimated. Further empirical research is needed to compare the realized response to selection from visual versus HTP-based selection methods.

HTP traits have been shown to improve the ability to predict GY in wheat through multi-trait GS (Rutkoski et al., 2016; Sun et al., 2017; Crain et al., 2018); however, the ability to apply this approach to the seed-limited stage of breeding programs has not been evaluated. In this study, GY and HTP records assessed in the YT were integrated with genomic information to train multi-trait GS models. These

were then applied to predict GY for the SP, where collecting HTP traits is possible but GY is not typically recorded. Overall, the multi-trait models showed the highest predictive abilities as compared to univariate GS and prediction with the HTP traits alone, though the increase over the latter was marginal. Therefore, it may be more cost-effective for breeding programs to invest only in HTP as opposed to both HTP and genotyping for GS. While the cost of genotyping has reduced greatly over the past decade, the large number of individuals in breeding programs at the early stages can make applying GS expensive, particularly for public sector programs which generally have smaller budgets. However, if genomic markers and HTP traits were both available for a given early stage breeding population, it would be optimal to use both sources of information to build prediction models.

Multiple findings from this study suggested that using NDVI as the selection criteria at early breeding stages would have driven marked changes in DTHD within the breeding population. The potential for undesired directional response to selection in confounded traits is not unique to HTP but is rather relevant to any selection scheme in which other traits are confounded with the trait of interest. For any novel selection method, it is important to consider how selection for particular criteria may affect the distributions of genetically correlated traits.

In wheat, GY and other stress-adaptive traits can often be confounded with phenology (Reynolds et al., 2009). Because the various phenological stages in wheat show differential sensitivity to the environment (Fischer, 1985), the direction of the relationship between GY and phenology can depend on the environmental conditions at the testing location. For example, Blum and Pnuel (1990) evaluated that relationship

across a rainfall gradient and found that, under low-yielding and dry conditions, the correlation was negative with early heading varieties having a relative yield advantage, while in high-yielding environments, varieties that reached heading later in the season had higher yields. A similar trend has been observed in the elite yield trial stage of the CIMMYT breeding program. Moderate negative correlations between DTHD and GY were observed for the heat- and drought-stressed treatments (Juliana et al., 2019), while the two traits showed a moderate positive correlation in the irrigated bed-planted treatment (Juliana et al., 2018). However, the relationship was inconsistent in strength and direction across years for the irrigated flat-planted treatment, in which lodging had a strong effect on GY.

The correlation between phenology and GY may also depend on the extent to which alleles conferring phenology are segregating in the breeding population. CIMMYT maintains variation for DTHD in its bread wheat breeding program in order to develop varieties for a wide range of latitudes. By comparison, for a regional breeding program, it is likely that the phenology of the program's germplasm resources will be narrowed, providing the best adaptation to the local environment. .

The association between HTP and phenology is likewise well documented. Since before the development of UAV platforms for agricultural research, VIs derived from satellites have been used to track phenology in wheat for commercial and agronomic uses (Benedetti and Rossini, 1993; Boken and Shaykewich, 2002). Several other stress-adaptive traits that can be measured with HTP systems are often confounded with phenology in wheat. Canopy temperature (CT), which indicates crop water status (Jackson et al., 1977), has shown to be useful for identifying genotypes

that are tolerant to heat and drought (Amani et al., 1996; Reynolds et al., 2009) and can be measured using infrared thermometers or cameras deployed on ground- (as in Andrade-Sanchez et al., 2014; Crain et al., 2017) or aerial-based (as in Rutkoski et al., 2014, Sun et al., 2017) HTP systems. However, Mason and Singh (2014) demonstrated that selection on CT for wheat GY under drought conditions may result in directional selection on phenology, due to significant confounding of CT and phenology.

In this study, GY, DTHD, and the NDVI traits were all highly positively correlated with one another in the positive direction. Therefore, while selecting on HTP at the SP stage of the breeding program would have resulted in a considerable positive response to selection in GY at the YT stage, the distribution of DTHD would have also shifted, particularly at higher levels of selection intensity, to favor later-maturity individuals. We presented two approaches to account for the confounding effects of phenology: 1) correcting the GY and HTP traits for phenology by including DTHD as a fixed effect covariate during prediction and 2) utilizing a restricted selection index to identify the top performing individuals for GY without selecting directionally on DTHD.

For the first approach, given that high correlations were observed between GY, DTHD, and HTP, the predictive abilities of HTP selection and the multi-trait GS models were greatly reduced following the correction for DTHD to the extent that the addition of HTP traits in multi-trait GS provided no additional predictive power compared to the univariate GS. Estimates of response to selection for GY in the YT were low, less than 1 percent for most HTP time-points, when using DTHD-corrected

HTP traits evaluated in the SP as the selection criteria.

By including DTHD as a fixed effect covariate during prediction, the genetic correlations between DTHD and GY or HTP are ignored. Any of the genetic variance that may have been contributing to both traits is partitioned into the fixed effect. A potential strategy for assessing the value of HTP for selection independently of the confounding effects of phenology would be to conduct a similar experiment on a population with limited phenological variation. Genetic mapping populations like the Seri/Babax recombinant inbred lines (Olivares-Villegas et al., 2007) and the wheat association mapping initiative diversity panel (Lopes et al., 2015) were designed to minimize the phenotypic variance for phenology traits. These populations were created to enable the identification of marker-trait associations in the absence of the confounding effects of phenology. Using the Seri/Babax population, Lopes and Reynolds (2012) demonstrated that estimates of NDVI at physiological maturity were both moderately heritable and independent of phenology. While these populations may be useful for assessing the ability of HTP to capture genetic variation underlying traits of interest in the absence of confounding, they do not reflect the phenological distributions of the CIMMYT bread wheat breeding program, among others. Therefore, breeding methods that can effectively account for the confounding relationships between HTP and phenology without losing useful information are needed.

In this study, data on visual selections made by the breeders in the SP_{BP} showed that, when performing visual selection, the breeders do not select directionally on DTHD. A restricted selection index can mimic this by enabling selection for

increased GY while holding DTHD fixed. This approach was utilized to compare the predictive abilities of HTP selection, univariate GS, and multi-trait GS to identify individuals with high GY potential across the distribution of DTHD.

The results from 2016-17 showed that HTP selection demonstrated similar levels of predictive ability as univariate GS when predicting the index, and that the multi-trait GS models that combined genomic marker information with HTP traits further increased accuracies. However, in 2017-18, HTP selection performed more poorly than univariate GS, and multi-trait models were only marginal better than the univariate case. The HTP traits in the SP were more strongly correlated with the index in 2016-17 than in 2017-18, and the index was more heritable in 2016-17 than in 2017-18.

Due to the possibility of unintended indirect selection on confounded traits, we recommend integrated approaches that leverage useful information from the HTP traits without relying on them fully. For example, van Ginkel et al. (2008) compared elite wheat breeding lines for GY under drought and heat stress developed through visual selection and through an integrated approach combining visual selection with measurements of canopy temperature depression. Although the lines with the highest GY for both methods were statistically similar, the integrated approach was more effective in identifying and eliminating the lower yielding lines. In this study, despite low correlations between HTP and GY in the SP_{BP}, the HTP traits were likewise more efficient than visual selection with respect to identifying low potential individuals. To avoid directional selection on phenology, we showed that HTP traits could be used in conjunction with qualitative DTHD scores to identify individuals with higher GY

potential for the “early,” “mid,” and “late” heading subgroups.

As HTP systems become more automated, customizable, and scalable, it is likely that breeders will have access to more extensive suites of phenotypic traits in addition to VIs. For example, hyperspectral reflectance phenotypes, which record spectral reflectance at a large range of wavelengths, were shown to increase prediction accuracy for GY when compared to VIs (Montesinos-López et al., 2017). Beyond spectral reflectance-related traits, deep learning algorithms have been developed to determine DTHD from proximal imagery of wheat canopies (Wang et al., 2019), UAV imagery has been used to estimate lodging in wheat as a function of changes in plant height throughout the growing season (Singh et al., 2019), and convolutional neural networks have been trained to identify foliar diseases in maize from aerial imagery (DeChant et al., 2017). These comprehensive suites of traits may allow confounding effects such as those observed in this study to be adequately addressed through integrative selection approaches that account for genetic correlations between traits.

Although recent advancements in remote sensing and robotic technologies have produced inexpensive cameras, sensors, UAV, smartphone applications, etc., the cost of deploying HTP platforms can be considerable (reviewed by Reynolds et al., 2019). Additional expenses may include the labor required to deploy HTP platforms in the field, training of personnel and dedicated bioinformaticians, the development and implementation of data extraction pipelines, propriety software solutions, and the establishment and maintenance of systems for storing and organizing large volumes of phenotypic data. Future improvements in the automation of HTP platforms and data analysis workflows as well as the development of open source software may help to

alleviate some of the costs of implementing HTP, though they are not likely to decrease below the costs associated with visual selection. Therefore, the breeder must determine if the added investment in HTP can be offset by increases in the rate of genetic gain and/or the costs saved from sowing fewer replicated yield trials.

Conclusion

We have demonstrated that aerial HTP traits collected at the seed-limited stage have the potential to improve selection strategies for GY in wheat. VIs measured on small, unreplicated plots were found to be heritable and correlated with GY in the replicated yield trials. The outlined approaches could be useful in breeding programs of crops that require one or more generations of seed increase before there is sufficient seed to sow replicated yield trials or at stages of the breeding program when the number of entries is too large to feasibly measure yield. Compared to GS, the VIs alone produced higher predictive abilities for GY, which may represent a cost-savings opportunity for breeding programs with insufficient resources to genotype the large number of individuals typically assessed at the early stages, though multi-trait GS models that integrated both genomic information and HTP traits gave the highest predictive abilities overall.

This study also showed that utilizing the VIs for indirect selection for GY would have also resulted in a considerable shift in the phenological distribution of the breeding population, particularly at high levels of selection intensity, due to confounding between the VIs, DTHD, and GY. In this situation, we recommend utilizing selection indices that are able to take into account the genetic correlations between traits, though results were mixed with respect to the ability of the VIs to

predict a restricted selection index developed to improve GY while holding DTHD fixed.

Although the results of a comparison between visual and HTP-based selection strategies were promising, further empirical studies are necessary to assess realized gains of each method. Several experimental design elements associated with the early generation seed-limited stage of breeding programs – including the absence of replication, confounding between genetic and spatial variation, and the reluctance of breeders to evaluate unselected material for GY – that make it difficult to develop reliable estimates of trait heritability and to draw comparisons between selection methods.

REFERENCES

Amani, I., R.A. Fischer, and M.P. Reynolds, 1996 Canopy temperature depression association with yield of irrigated spring wheat cultivars in a hot climate. J. Agron. Crop Sci. 176: 119-129.

Andrade-Sanchez, P., M.A. Gore, J.T. Heun, K.R. Thorp, A.E. Carmo-Silva, A.N. French, M.E. Salvucci, and J.W. White, 2014 Development and evaluation of a field-based high-throughput phenotyping platform. Func. Plant Biol. 41: 68-79.

Appels, R., Eversole, K., Feuillet, C., Keller, B., Rogers, J., Stein, N., Pozniak, C.J., Choulet, F., Distelfeld, A., and J. Poland, 2018 Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science 361: eaar7191.

Araus, J.L. and J.E. Cairns, 2014 Field high-throughput phenotyping: the new crop breeding frontier. Trends Plant Sci. 19: 52-61.

Babar, M.A., M.P. Reynolds, M. van Ginkel, A.R. Klatt, W.R. Raun, and M.L. Stone, 2006 Spectral reflectance indices as a potential indirect selection criteria for wheat yield under irrigation. Crop Sci. 46: 578-588.

Benedetti, R., and P. Rossini, 1993 On the use of NDVI profiles as a tool for agricultural statistics: the case study of wheat yield estimate and forecast in Emilia Romagna. Remote Sens. Environ. 45: 311-326.

Bernardo, R., 2003 On the effectiveness of early generation selection in self-pollinated crops. Crop Sci. 43: 1558-1560.

Blum, A., and Y. Pnuel, 1990 Physiological attributes associated with drought resistance of wheat cultivars in a Mediterranean environment. Aust. J. Agric.

Res. 41: 799-810.

Boken, V.N., and C.F. Shaykewich, 2002 Improving an operational wheat yield model using phenological phased-based Normalized Difference Vegetation Index. Int. J. Remote Sens. 23: 4155-4168.

Borghi, B., M. Accerbi, and M. Corbellini, 1998 Response to early generation selection for grain yield and harvest index in bread wheat (*T. aestivum* L.). Plant Breeding 117: 13-18.

Brennan, J.P., 1988 *An economic investigation of wheat breeding programmes*. Agricultural Economics Bulletin 35, Department of Agricultural Economics and Business Management, University of New England, Armidale, NSW, Australia.

Cabrera-Bosquet, L., J. Crossa, J. von Zitzewitz, M.D. Serret, and J.L. Araus, 2012 High-throughput phenotyping and genomic selection: the frontiers of crop breeding converge. J. Integr. Plant Biol. 54: 312-230.

Clarke, F.R., R.J. Baker, and R.M. DePauw, 1998 Interplot interference distorts yield estimates in spring wheat. Crop Sci. 38: 62-66.

Crain, J.L., Y. Wei, J. Barker, S.M. Thompson, P.D. Alderman, M. Reynolds, N. Zhang, and J. Poland, 2016 Development and deployment of a portable field phenotyping platform. Crop Sci. 56: 965-975.

Crain, J., M. Reynolds, and J. Poland, 2017 Utilizing high-throughput phenotypic data for improved phenotypic selection of stress-adaptive traits in wheat. Crop Sci. 57: 648-659.

Crain, J., S. Mondal, J. Rutkoski, R.P. Singh, and J. Poland, 2018 Combining high-

throughput phenotyping and genomic information to increase prediction and selection accuracy in wheat breeding. The Plant Genome 11.

DeChant, C., T. Wiesner-Hanks, X. Chen, E.L. Stewart, J. Yosinski, M.A. Gore, R.J. Nelson, and H. Lipson, 2017 Automated identification of Northern Leaf Blight-infected maize plants form field imagery using deep learning. Phytopathology 107: 1426-1432.

Dornbusch, T., R. Baccar, J. Watt, J. Hillier, J. Bertheloot, C. Fournier, and B. Andrieu, 2011 Plasticity of winter wheat modulated by sowing date, plant population density and nitrogen fertilisation: dimensions and size of leaf blades, sheathes and internodes in relation to their position on a stem. Field Crops Res. 121: 116-124.

Edmeades, G.O., and T.B. Daynard, 1979 The development of plant-to-plant variability in maize at different planting densities. Can. J. Plant Sci. 59: 561-576.

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, and S.E. Mitchell, 2011 A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PloS ONE 6: e19379.

Endelman, J.B., 2011 Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Genome 4: 250-255.

Endelman, J.B., and J.L. Jannink, 2012 Shrinkage estimation of the realized relationship matrix. G3 Genes Genom. Genet. 2: 1405-1413.

Fischer, R.A., 1985 Number of kernels in wheat crops and the influence of solar radiation and temperature. J. Agr. Sci. 105: 447-461.

Fonseca, S., and F.L. Patterson, 1968 Yield component heritabilities and interrelationships in winter wheat (*Triticum aestivum* L.). Crop Sci. 8: 614-617.

Garrick, D.J., Taylor, J.F., and R.L. Fernando, 2009 Deregressing estimated breeding values and weighting information for genomic regression analyses. Genet. Sel. Evol. 41: 55.

Gilmour, A.R., B.J. Gogel, B.R. Cullis, S.J. Welham, and R. Thompson, 2014 ASReml user guide release 4.1 structural specification. VSN International, Hemel Hempstead, UK.

Glaubitz, J.C., Casstevens, T.M., Lu, F., Harriman, J., Elshire, R.J., Sun, Q., and E.S. Buckler, 2014 TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. PloS ONE 9: e90346.

Gu X.F., Guyot G, and M. Verbrugghe, 1992 Evaluation of measurement errors in ground surface reflectance for satellite calibration. Int. J. Remote Sense. 13: 2531-2546.

Haghighattalab, A., L. González-Pérez, S. Mondal, D. Singh, D. Schinstock, J. Rutkoski, I. Ortiz-Monasterio, R.P. Singh, D. Goodin, and J. Poland 2016 Application of unmanned aerial systems for high throughput phenotyping of large wheat breeding nurseries. Plant Methods 12:35.

Hazel, L.N., 1943 The genetic basis for constructing selection indexes. Genetics 28: 476-490.

Huete, A., K. Didan, T. Miura, E.P. Rodriguez, X. Gao, and L.G. Ferreria, 2000 Overview of the radiometric and biophysical performance of the MODIS

vegetation indices. Remote Sens. Environ. 83: 195-213.

Jackson, R.D., R.J. Reginato, and S.B. Idso, 1977 Wheat canopy temperature: a practical tool for evaluating water requirements. Water Resour. Res. 13: 651-656.

Joshi, A.K., O. Ferrara, J. Crossa, G. Singh, R. Sharma, R. Chand, and R. Parsad, 2007 Combining superior agronomic performance and terminal heat tolerance with resistance to spot blotch (*Bipolaris sorokiniana*) in the warm humid Gangetic plains of south Asia. Field Crops Res. 103: 53-61

Juliana, P., R.P. Singh, J. Poland, S. Mondal, J. Crossa, O.A. Montesinos-López, S. Dreisigacker, P. Pérez-Rodríguez, J. Huerta-Espino, L. Crespo-Herrera, and V. Govindan, 2018 Prospects and challenges of applied genomic selection – a new paradigm in breeding for grain yield in bread wheat. The Plant Genome.

Juliana, P., O.A. Montesinos-López, J. Crossa, S. Mondal, L. González Pérez, J. Poland, J. Huerta-Espino, L. Crepso-Herrera, V. Govindan, S. Dreisigacker, S. Shrestha, P. Pérez-Rodríguez, F. Pinto Espinosa, and R.P. Singh, 2019 Integrating genomic-enabled prediction and high-throughput phenotyping in breeding for climate-resilient bread wheat. Theor. Appl. Genet. 132: 177-194.

Kempthorne, O., and A.W. Nordskog, 1959 Restricted selection indices. Biometrics 15: 10-19.

Knott, D.R., and J. Kumar, 1975 Comparison of early generation yield testing and a single seed descent procedure in wheat breeding. Crop Sci. 15: 295-299.

Krause, M.R., L. González-Pérez, J. Crossa, P. Pérez-Rodríguez, O. Montesinos-López, R.P. Singh, S. Dreisigacker, J. Poland, J. Rutkoski, M. Sorrells, M.A.

Gore, and S. Mondal, 2019 Hyperspectral reflectance-derived relationship matrices for genomic prediction of grain yield in wheat. G3 Genes Genom. Genet. 9: 1231-1247.

Langmead, B., and S.L. Salzberg, 2012 Fast gapped-read alignment with Bowtie 2. Nature Methods 9: 357-359.

Lopes, M.S., and M.P. Reynolds, 2012 Stay-green in spring wheat can be determined by spectral reflectance measurements (normalized difference vegetation index) independently from phenology. J. Exp. Bot. 63: 3789-3798.

Lopes, M.S., S. Dreisigacker, R.J. Peña, S. Sukumaran, and M.P. Reynolds, 2015 Genetic characterization of the wheat association mapping initiative (WAMI) panel for dissection of complex traits in spring wheat. Theor. Appl. Genet. 128: 435-464.

Magnussen, S., 1993 Bias in genetic variance estimates due to spatial autocorrelation. Theor. Appl. Genet. 86: 349-355.

Mason, R.E., and R.P. Singh, 2014 Considerations when deploying canopy temperature to select high yielding wheat breeding lines under drought and heat stress. Agronomy 4: 191-201.

McLaren, C.G., L. Ramos, C. Lopez, and W. Eusebio, 2000 Application of the genealogy management system, pp. 5.8-5.13 in *International Crop Information System. Technical Development Manual, version VI,* edited by C.G. McLaren, J.W. White, and P.N. Fox. CIMMyT, México: CIMMyT and IRRI.

Mitchell, J.W., R.J. Baker, and D.R. Knott, 1982 Evaluation of honeycomb selection for single plant yield in durum wheat. Crop Sci. 22: 840-843.

Mondal, S., R.P. Singh, J. Crossa, J. Huerta-Espino, I. Sharma, R. Chatrath, G.P. Singh, V.S. Sohu, G.S. Mavi, V.S.P. Sukuru, I.K. Kalappanavar, V.K. Mishra, M. Hussain, N.R. Gautam, J. Uddin, N.C.D. Barma, A. Hakim, and A.K. Joshi, 2013 Earliness in wheat: a key adaptation under terminal and continual high temperature stress in South Asia. Field Crops Res. 151: 19-26.

Montesinos-López, O.A., A. Montesinos-López, J. Crossa, G. de los Campos, G. Alvarado, M. Mondal, J. Rutkoski, L. González-Pérez, and J. Burgueño, 2017 Predicting grain yield using canopy hyperspectral reflectance in wheat breeding data. Plant Methods 13: 4.

Olivares-Villegas, J.J., M.P. Reynolds, and G.K. McDonald, 2007 Drought-adaptive attributes in the Seri/Babax hexaploid wheat population. Funct. Plant Biol. 34: 189-203.

Openshaw, S.J., and H.H. Hadley, 1984 Selection indexes to modify protein concentration of soybean seeds. Crop Sci. 24: 1-4.

Pauli, D., S.C. Chapman, R. Bart, C.N. Topp, C.J. Lawrence-Dill, J. Poland, and M.A. Gore, 2016 The quest for understanding phenotypic variation via integrated approaches in the field environment. Plant Physiol. 172: 622-634.

Pešek, J., and R.J. Baker, 1969 Desired improvement in relation to selection indices. Can. J. Plant Sci. 49: 803-804.

Poland, J., Endelman, J., Dawson, J., Rutkoski, J., Wu, S., Manes, Y., Dreisigacker, S., Crossa, J., Sánchez-Villeda, H., and M. Sorrells, 2012 Genomic selection in wheat breeding using genotyping-by-sequencing. The Plant Genome 5: 103-113.

Prasad, B., B.F. Carver, M.L. Stone, M.A. Babar, W.R. Raun, and A.R. Klatt, 2007 Potential use of spectral reflectance indices as a selection tool for grain yield in winter wheat under Great Plains conditions. Crop Sci. 47: 1426-1440.

Prasanna, B.M., J.L. Araus, J. Crossa, J.E. Cairns, N. Palacios, B. Das, and C. Magorokosho, 2013 High-throughput and precision phenotyping for cereal breeding programs. *In* P.K. Gupta and R.K. Varshney (eds.) Cereal genomics II. Springer, p. 314-374.

Quail, K.J., R.A. Fischer, and J.T. Woods, 1989 Early generation selection in wheat. I. Yield potential. Aust. J. Agric. Res. 40: 1117-1133.

R Core Team, 2016 R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rebetzke, G.J., A.G. Condon, R.A. Richards and G.D. Farquhar, 2002 Selection for reduced carbon isotope discrimination increases aerial biomass and grain yield of rainfed bread wheat. Crop Sci. 42: 739-745.

Rebetzke, G.J., Fischer, R.T.A., van Herwaarden, A.F., Bonnet, D.G., Chenu, K., Rattey, and N.A. Fettell, 2014 Plot size matters: interference from intergenotypic competition in plant phenotyping studies. Func. Plant Bio. 41: 107-118.

Reynolds, D., F. Baret, C. Welcker, A. Bostrom, J. Ball, F. Cellini, A. Lorence, A. Chawade, M. Khafif, K. Noshita, M. Mueller-Linow, J. Zhou, and F. Tardieu, 2019 What is the cost of phenotyping? Optimizing costs for different scenarios. Plant Sci. 282: 14-22.

Reynolds, M.P., Y. Manes, A. Izanloo, and P. Langridge, 2009 Phenotyping for physiological breeding and gene discovery in wheat. Ann. Appl. Biol. 155: 309-320.

Reynolds, M., and P. Langridge, 2016 Physiological breeding. Curr. Opin. Plant Biol. 31: 162-171.

Rutkoski, J.E., J. Poland, S. Mondal, E. Autrique, L. González Pérez, J. Crossa, M. Reynolds, and R. Singh, 2016 Canopy temperature and vegetation indices from high-throughput phenotpying improve accuracy of pedigree and genomic selection for grain yield in wheat. G3 Genes Genom. Genet. 6: 2799-2808.

Singh, D., X. Wang, U. Kumar, L. Gao, M. Noor, M. Imtiaz, R.P. Singh, and J. Poland, 2019, High-throughput phenotyping enabled genetic dissection of crop lodging in wheat. Front. Plant Sci. 10: 394.

Singh, R.P., S. Rajaram, A. Miranda, J. Huerta-Espinoza, and E. Autrique, 1998 Comparison of two crossing and four selection schemes for yield, yield traits, and slow rusting resistance to leaf rust in wheat. Euphytica 100: 35-43.

Smith, H.F., 1936 A discriminant function for plant selection. Ann. Eugen. 7: 240-250.

Snape, J.W. and E. Simpson, 1984 Early generation selection and rapid generation advancement methods in autogamous crops. *In* W. Lange et al., (eds.) Efficiency in plant breeding. Proc. 10th Cong. EUCARPIA. Pudoc, Wageningen, The Netherlands.

Sun, J., J.E. Rutkoski, J.A. Poland, J. Crossa, J.L. Jannink, and M.E. Sorrells, 2017 Multitrait, random regression, or simple repeatability model in high-throughput

phenotyping data improve genomic prediction for wheat grain yield. The Plant Genome 10.

Syme, J.R., 1972 Single-plant characters as a measure of field plot performance of wheat cultivars. Aust. J. Agric. Res. 23: 753-760.

van Ginkel, M., R.M. Trethowan, K. Ammar, J. Wang, M. Lillemo, 2002 Guide to bread wheat breeding at CIMMYT. Wheat Program Special Report No. 5. International Maize and Wheat Improvement Centre, Mexico, D.F., Mexico.

van Ginkel, M., M.P. Reynolds, R. Trethowan, and E. Hernandez, 2008 Complementing the breeder's eye with canopy temperature measurements. *In* M.P. Reynolds, J. Pietragalla, and J.H. Braun (eds.) International symposium on wheat yield potential: challenges to international wheat breeding. CIMMYT: Edo. de México, Mexico.

Wang, X., H. Xuan, B. Evers, S. Shrestha, R. Pless, and J. Poland, 2019 High-throughput phenotyping with deep learning gives insights into the genetic architecture of flowering time in wheat. bioRxiv: 527911.

Watanabe, K., W. Guo, K. Arai, H. Takanashi, H. Kajiya-Kanegae, M. Kobayashi, K. Yano, T. Tokunaga, T. Fujiwara, N. Tsutsumi, and H. Iwata, 2017 High-throughput phenotyping of sorghum plant height using an unmanned aerial vehicle and its application to genomic prediction modeling. Front. Plant Sci. 8: 421.

White, J., and J.L. Harper, 1970 Correlated changes in plant size and number in plant populations. J. Ecol. 58: 467-485.

CHAPTER 3

A RANDOM FOREST REGRESSION-BASED APPROACH FOR OPTIMIZING VARIABLE PLANTING RATES FOR CORN AND SOYBEAN USING HIGH-RESOLUTION TOPOGRAPHICAL AND SOIL DATA

Abstract

In recent years, planting machinery that enables precise control of the planting rates has become available for corn (*Zea mays* L.) and soybean (*Glycine max* L.). Combined with increasingly available topographical and soil information, there has been a growing interest in developing planting rate designs to exploit variation in the agri-landscape in order to maximize production. A random forest regression-based approach was developed to model the interactions between planting rate, topography, and soil characteristics and their effects on yield based on on-farm variable rate planting trials for corn and soybean conducted at 27 locations in New York between 2014 and 2018 through a collaborative project with the New York Corn and Soybean Growers Association. Planting rate ranked highly in terms of random forest regression variable importance, despite explaining relatively low yield variation in the ANOVA, suggesting that yield response to planting rate likely depends on underlying agri-landscape features. The method was extended to identify the optimal planting rates for maximizing yield or profit given the underlying topographical and soil data. When maximizing for yield, the resulting optimized designs suggested that variable rate planting may be more advantageous for corn, while fixed rate sowing at high densities

may be preferable for soybean. When maximizing for profit for corn, variable rate planting appeared to be advantageous only when the seed cost to crop market price ratio is favorable to growers, suggesting that the marginal yield increases from sowing higher rates do not offset the additional cost of seed at average or unfavorable cost-to-price ratios. For soybean, the resulting optimized designs were inconsistent across site-years, with some site-years assigned to variable rates while others were predominantly allocated fixed rates, suggesting that the marginal benefit of planting at higher seed rates is largely site-year-dependent. Additionally, yield predictions across site-years were relatively low. Together, these results suggest that local testing may provide more accurate optimized planting rate designs. Further long-term studies are needed to assess the stability of optimal planting rate designs and to validate them empirically.

Introduction

In agricultural production systems, site-specific management involves the development of crop management strategies at a finer spatial scale than that of the whole field area (Plant, 2001). Compared to early farming, which was performed by hand and thereby facilitated site-specific management, the advent of mechanization in farming ushered in an era of the uniform application of inputs on larger areas of cropland. Studies have shown that uniform applications can result in suboptimal input use efficiency in parts of the field that may require greater or fewer inputs than the applied fixed rated (Mulla and Schepers, 1997; Moore and Tyndale-Biscoe, 1999), though the associated economic losses were generally outweighed by the increases in farm productivity. However, the growing costs of agricultural inputs (USDA NASS, 2017) and negative environmental effects of intensive production practices (Tilman et

al., 2002) threaten the long-term economic and environmental sustainability of uniform crop management.

Variable rate application systems enable growers to apply inputs at any number of user-defined levels within a field area. Variable rate as it applies to planting is not a new concept. The first variable rate planting systems emerged in the United States during the 1970s but did not become popular due to their initial lack of automation, which necessitated that growers trigger planting rate transitions manually (Lowenberg-DeBoer, 1999). The integration of GPS and automated variable rate systems for commercial planting equipment renewed interest in variable rate planting technologies in the late 1990s.

However, the improvements in equipment and technology did not immediately drive a widespread adoption of variable rate planting, as users remained uncertain of how to develop variable rate planting designs given the underlying conditions of their fields. Some early and more recent studies have based the development of variable rate planting designs on yield potential, delineated by previous years' yield performance and/or the grower's knowledge of the field (Barnhisel et al., 1996; Hörbe et al., 2013; Corassa et al., 2018). However, Bullock et al., (1998) postulated that knowledge of yield potential would be insufficient to make variable rate technologies economically beneficial to farmers. They identified as a major barrier the community's limited understanding of how planting rate and various field characteristics such as topography and soil – which can be expensive to measure – may interact to influence yield. More than 20 years later, despite the increasing availability of high-resolution topographical and soil data and the abundance of

commercial variable rate products on the market, the use of the variable rate planting technology remains limited. In a 2016 survey of New York State corn and soybean growers, only 10.5 percent of respondents had adopted variable rate planting technologies, and 20.0 percent cited skepticism regarding the economic benefits as a driver of the technology's slow adoption (van Es et al., 2016).

Currently there is little consensus among the community regarding how to create variable rate planting designs with respect to yield potential, field characteristics, and/or other observations. While many research efforts have been undertaken to delineate the relationships between topographic features, soil characteristics, and yield potential for crops sown under fixed rates (Katerji et al., 1995; Miller et al., 1988; Kravchenko and Bullock, 2000; Frogbrook et al., 2002; Cox et al., 2003; Changere and Lal, 2008), very few studies have assessed how planting rate might effect and interact with those relationships. To date, the results have been mixed: some works have observed potential for the profitability for variable rate planting technology (Shanahan et al., 2004), while others found inconsistencies among planting rate optimizations and the interactions between planting rate and topographical/soil features across sites and years (Smidt et al., 2016; Licht et al., 2017).

In a literature review of variable rate technologies, Bullock and Lowenberg-DeBoer (2007) made the observation that the majority of studies pertaining to variable rate planting and chemical applications have reported results for a small number of specific site-years. The authors, among others (Lambert et al., 2006; Liu et al., 2006; Ruffo et al., 2006; Bullock et al., 2007), emphasized the need for longer-term, multi-

location experimental data in order to expand the inference space and sample the population of possible environmental conditions at a high level in order to evaluate the potential for variable rate planting.

To conduct experiments on such a scale, coordination between many growers and researchers are needed in addition to a long-term commitment to testing. The New York Corn and Soybean Growers Association (NYCSGA) is a statewide non-profit organization that represents the interests of New York's corn and soybean producers. Among other activities, the association sponsors relevant research on corn and soybean production, utilization and marketing. In 2013, the NYCSGA developed a research initiative to optimize variable rate planting technology for corn and soybean growers in New York State. NYCSGA members throughout central, western, and northern New York have conducted on-farm variable rate field trials between 2014 and 2018 with the objective of developing a strategy for variable rate planting to exploit the extensive native variation of New York's agricultural landscape. To capture this variation, spatial data related to topographical features, soil type, and soil nutrients were collected at a high resolution.

This collaborative project driven by growers has created a wealth of information relating yield to planting rate and its interactions with a wide range of environmental factors. However, from a statistical standpoint, a number of challenges exist when analyzing these types of agronomic data that may result in a loss of information or cause problems with the interpretation of results. To date, most of the published studies evaluating the potential of variable rate planting technology have relied on multiple linear regression to assess the relationships between planting rate,

yield, and environmental features (Shanahan et al., 2004; Hörbe et al., 2013; Licht et al., 2017). However, linear modeling possesses a number of weaknesses in this context. First, linear regression methods handle multicollinearity between predictor variables relatively poorly. Many oft-recorded variables in environmental or agricultural studies exhibit significant correlations among one another. For example, topographical attributes such as elevation and slope can influence the movement of water through and over the landscape, thereby affecting the development of soil (Moore et al., 1993). Relationships between these topographical features and soil characteristics such as sand, silt, organic matter, pH, cation exchange capacity, and extractable phosphorus, among others, have been observed (Brubaker et al., 1993; Moore et al., 1993; Tan et al., 2004). Likewise, strong relationships between soil variables themselves are common. In a linear-regression analysis of the effects of soil characteristics on corn yield response to variable planting rates, Licht et al., (2017) removed variables describing silt and available water-holding capacity due to their collinearity with sand, clay, and soil organic matter. Multicollineary between variables can cause problems with model fit and interpretation; though removing correlated but imperfectly correlated characters may result in a loss of information. In addition to correlations between variables, nonlinear relationships can often occur in agricultural systems (reviewed by Archontoulis and Miguez, 2014). For example, increasing temperatures are known to benefit corn yields but become harmful once they reach 30°C and above (Schlenker and Roberts, 2006), a relationship for which linear approaches would model poorly.

Novel statistical approaches to analyze these increasingly large and complex

agronomic datasets are needed to fully exploit the information they contain. Random forest regression is an ensemble learning method that constructs multiple decision trees using random subsets of the observations and predictor variables (Breiman, 2001). Random forest regression has become increasingly more popular in a range of scientific fields due to its ability to account for highly correlated predictor variables and model complex nonlinear interactions. Random forest regression has been used effectively for environmental and agricultural applications such as predicting soil texture from terrain variables (Ließ et al., 2012) and estimating vegetative biomass from remotely sensed data (Mutanga et al., 2012). This method may also be useful for understanding the complex relationships between planting rate, yield, and field characteristics and for developing optimized variable rate planting designs.

The work presented here aims to 1) discuss the interactions of planting rate with topographical and soil features and their subsequent effects on yield, 2) propose a random forest regression-based method for leveraging this information to build variable rate planting designs, and 3) discuss their potential for use and profitability.

Materials and Methods

Site-Year Summary

On-farm corn and soybean variable rate field trials were conducted in 27 unique fields throughout central, western, and northern New York (Figure 3.1) between the years 2014 and 2018 for a total of 57 site-years (32 corn, 25 soybean). Site-year-specific information detailing field locations, areas, row spacings, planting rates and hybrids/varieties sown, and experimental designs used can be found in Table 3.1. Field areas ranged from 10.8 to 59.9 ha with an average of 28.0 ha. The row

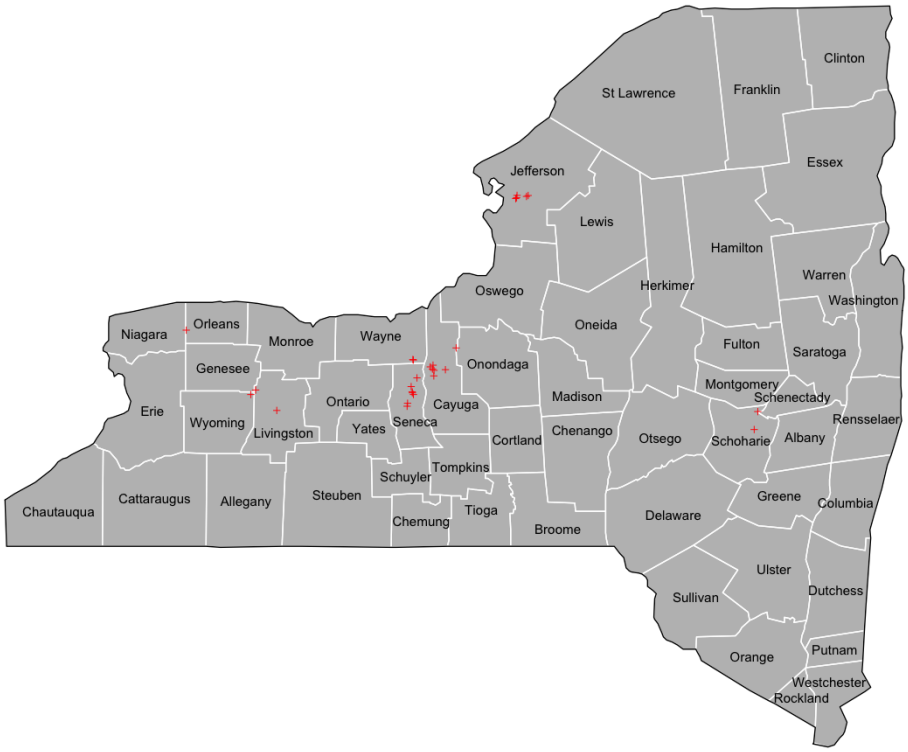


Figure 3.1 County-level map of the on-farm field trials conducted between 2014 and 2018 in New York

Table 3.1: Site-year-specific information for each on-farm field trial

Field	Year	Crop	Area (ha)	Row Spacing (cm)	Hybrid/Variety	Planting Rate Design	Hybrid/Variety Design
An6	2014	Corn	22.81	50.8	P9690AM P9675AMXT	Randomized Blocks	Alternating Hybrids/Varieties
An8	2014	Corn	29.99	50.8	P0216 P0533AM1	Randomized Blocks	Alternating Hybrids/Varieties
An8	2015	Corn	29.99	50.8	P0216 P0533AM1	Randomized Blocks	Alternating Hybrids/Varieties
Ba1	2015	Corn	21.23	76.2	P0216 P0533AM1	Randomized Blocks	Alternating Hybrids/Varieties
Be2	2014	Corn	51.15	76.2	P0216 P0533AM1	Randomized Blocks	Alternating Hybrids/Varieties
Bi	2015	Corn	14.92	101.6	P0533AM1	Randomized Blocks	Single Hybrid/Variety
Bi	2016	Corn	14.92	101.6	P0157AMX	Randomized Blocks	Single Hybrid/Variety
Bi	2016	Corn	14.92	101.6	197-68STXRIB	Randomized Blocks	Single Hybrid/Variety
Cr	2014	Corn	29.55	101.6	DKC46-20RIB P0094AMX	Randomized Blocks	Alternating Hybrids/Varieties
Cr	2015	Corn	29.55	101.6	P0216AM	Randomized Blocks	Single Hybrid/Variety
Dm2	2014	Corn	23.47	50.8	P0216AM P0533AM1	Randomized Blocks	Alternating Hybrids/Varieties
Du1	2014	Corn	24.34	76.2	P0216AM P0533AM1	Randomized Blocks	Alternating Hybrids/Varieties
Du3	2016	Corn	39.63	76.2	P0216AM P0533XR	Randomized Blocks	Alternating Hybrids/Varieties
He15	2015	Corn	23.75	50.8	P0157AMX DK52-85	Randomized Blocks	Alternating Hybrids/Varieties
Hk	2015	Corn	35.07	50.8	P0533 P0604	Randomized Blocks	Alternating Hybrids/Varieties
Ke	2015	Corn	31.31	76.2	P0216AM P0533AM1	Randomized Blocks	Alternating Hybrids/Varieties
Mc2	2015	Corn	15.83	50.8	P0604 P0533AM1	Randomized Blocks	Alternating Hybrids/Varieties
Mc3	2015	Corn	17.78	50.8	P0506AM P0533AM1	Randomized Blocks	Alternating Hybrids/Varieties
Ov29	2014	Corn	23.97	101.6	P0094AMX	Randomized Blocks	Single Hybrid/Variety
Ov29	2015	Corn	23.97	101.6	P0094AMX	Randomized Blocks	Single Hybrid/Variety
Ov30	2014	Corn	10.80	101.6	P0094AMX	Randomized Blocks	Single Hybrid/Variety
Ov30	2015	Corn	10.80	101.6	P0604AM	Randomized Blocks	Single Hybrid/Variety
Ri23	2015	Corn	40.71	76.2	P0216AM P0533AM1	Randomized Blocks	Alternating Hybrids/Varieties
Ri27	2015	Corn	12.74	50.8	P0216AM P0533AM1	Randomized Blocks	Alternating Hybrids/Varieties
St	2015	Corn	47.71	76.2	P0216AM P0533AM1	Randomized Blocks	Alternating Hybrids/Varieties
Sc	2015	Corn	23.71	101.6	P0157AMX	Randomized Blocks	Single Hybrid/Variety
Sc	2016	Corn	23.71	101.6	P0533AM1	Randomized Blocks	Single Hybrid/Variety
Sc	2017	Corn	23.71	101.6	P0157AMX	Randomized Blocks	Single Hybrid/Variety

Table 3.1 (Continued)

Sw37	2015	Corn	23.72	76.2	49-72	Randomized Blocks	Single Hybrid/Variety
Sw52	2016	Corn	19.07	76.2	49-72	Randomized Blocks	Single Hybrid/Variety
Sw52	2017	Corn	19.07	76.2	49-72	Randomized Blocks	Single Hybrid/Variety
Vc1	2016	Corn	59.93	50.8	P0506AM P0533AM1 P0157AMX	Randomized Blocks	Alternating Hybrids/Varieties
An6	2015	Soybean	22.81	101.6	2108R	Randomized Blocks	Single Hybrid/Variety
Ba1	2016	Soybean	21.23	101.6	AG17X6	Alternating Strips	Single Hybrid/Variety
Do	2015	Soybean	56.64	101.6	AG3034	Randomized Blocks	Single Hybrid/Variety
Do	2016	Soybean	56.64	101.6	AG3334	Randomized Blocks	Single Hybrid/Variety
Du1	2015	Soybean	24.34	101.6	P24T19R	Alternating Strips	Single Hybrid/Variety
Du3	2015	Soybean	39.63	101.6	P24T19R 92Y91	Alternating Strips	Sections
He15	2014	Soybean	23.75	101.6	P92Y12 P92Y51	Randomized Blocks	Alternating Hybrids/Varieties
He15	2016	Soybean	23.75	101.6	P24T05	Randomized Blocks	Single Hybrid/Variety
Hk	2014	Soybean	35.07	127	P92Y51	Randomized Blocks	Single Hybrid/Variety
Ke	2016	Soybean	31.31	101.6	P19T01R	Alternating Strips	Single Hybrid/Variety
Lb	2016	Soybean	28.53	50.8	P19T/P01T P19T04R 15	Randomized Blocks	Sections
Mc1	2014	Soybean	25.17	101.6	AG3030	Randomized Blocks	Single Hybrid/Variety
Mc1	2018	Soybean	25.17	101.6	AG3334 SG1776	Randomized Blocks	Sections
Mc2	2014	Soybean	15.83	101.6	P93Y22	Randomized Blocks	Single Hybrid/Variety
Mc2	2016	Soybean	15.83	101.6	AG2035	Randomized Blocks	Single Hybrid/Variety
Mc3	2014	Soybean	17.78	101.6	AG3030	Randomized Blocks	Single Hybrid/Variety
Ov29	2018	Soybean	23.97	76.2	SG1863XT P16A35X	Randomized Blocks	Sections
Ov30	2018	Soybean	10.80	76.2	SG1863XT P16A35X	Randomized Blocks	Sections
St	2014	Soybean	47.71	101.6	P92Y51	Alternating Strips	Single Hybrid/Variety
Sw37	2014	Soybean	23.72	76.2	2031	Randomized Blocks	Single Hybrid/Variety
Sw37	2016	Soybean	23.72	76.2	AG2031	Randomized Blocks	Single Hybrid/Variety
Sw52	2015	Soybean	19.07	76.2	AG2431	Randomized Blocks	Single Hybrid/Variety
Tp80	2016	Soybean	31.27	76.2	Unknown	Randomized Blocks	Single Hybrid/Variety
Vc1	2015	Soybean	59.93	101.6	P22T41 P92Y51	Randomized Blocks	Sections
Vc1	2017	Soybean	59.93	101.6	P22T41 AG2035	Randomized Blocks	Sections

spacings used were 50.8, 76.2, 101.6, and 127.0 cm. For corn, 16 different hybrids were used throughout the course of the experiment, though hybrids P0533AM1 and P0216/P0216AM were used most predominantly, appearing in 15 and 8 site-years, respectively. For soybean, 21 varieties were used. The soybean variety used in site-year Tp80_2016 was unknown. Statewide weather data were obtained by the National Climatic Data Center (National Climatic Data Center, 2019). Monthly departures from the 1981-2010 averages for temperature and precipitation are illustrated in Figure 3.2.

On-Farm Experimental Designs

The field designs used for planting rates and hybrids/varieties are illustrated in Figure 3.3. For corn, the following four target planting rates were tested: 66,718 (27,000), 79,074 (32,000), 91,429 (37,000), and 103,784 (42,000) seeds ha^{-1} (ac^{-1}). All on-farm trials for corn were sown with a randomized design consisting of 0.81 ha (2 ac) target planting rate blocks. Out of the 32 site-years for corn, 17 were sown with two hybrids and 1 was sown with three hybrids. All on-farm trials with multiple hybrids were sown with a split planter, creating alternating strips of each hybrid within the field. The remaining site-years were sown with a single hybrid.

For soybean, the same 0.81 ha target planting rate block system was used. The soybean trials were sown with the following four target planting rates: 197,684 (80,000), 296,526 (120,000), 395,368 (160,000), and 494,210 (200,000) seeds ha^{-1} (ac^{-1}). Rates were assigned to blocks at random for 20 of the site-years. Because variable rate soybean planters were not available for the remaining 5 site-years, target planting rates were assigned to the rows or columns of the blocks, enabling adjustments to be made to the planting rate between passes of the planter. A total of 7 of the 25 site-

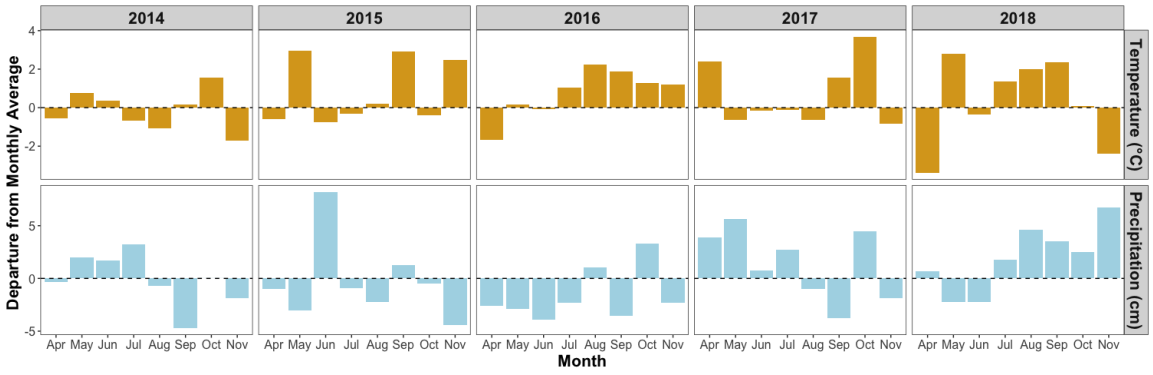


Figure 3.2: Statewide departures from 1981-2010 monthly temperature and precipitation averages during the 2014-2018 growing seasons

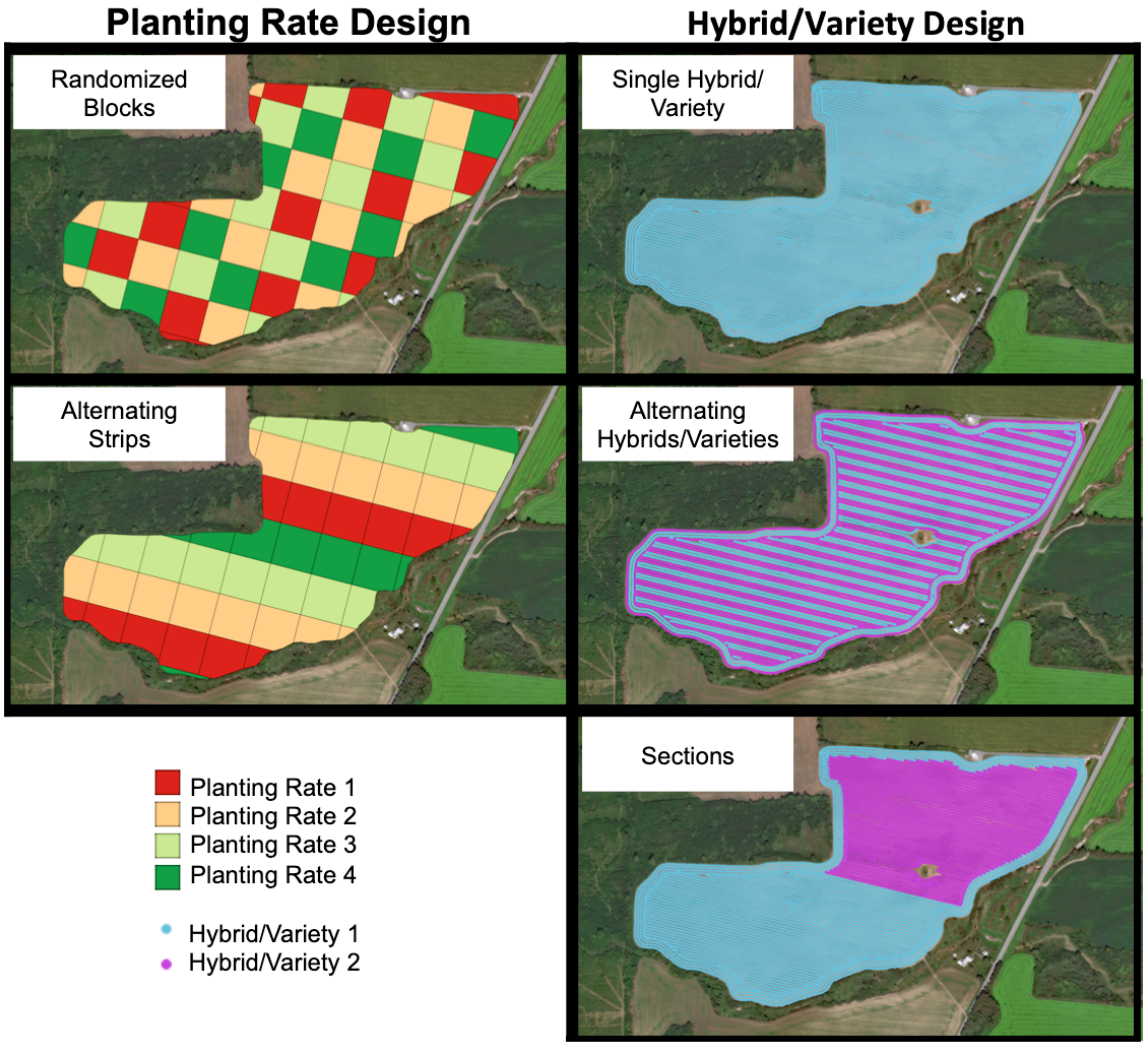


Figure 3.3: Field designs for planting rate and hybrid/variety

years were sown with two soybean varieties. A split soybean planter was only available for site-year He15_2014, which was sown with alternating varieties. The remaining 7 site-years were sown such that each variety occupied a section of the field.

Experimental Unit Grid

The 0.81 ha target planting rate block system was developed to minimize error associated with transitioning between planting rates. There are, however, limitations to treating the 0.81 ha target planting rate blocks as the experimental units. First, blocks that fell along the edges of the field were often irregularly shaped (seen in Figure 3.3). Secondly, the number of full-sized 0.81 ha blocks was low in smaller fields. Furthermore, due to the high variability of the topographical and soil features observed in some fields, using the blocks as the experimental units would require averaging over a potentially wide range of variability. Lastly, because split planters were used to sow multiple hybrids/varieties in some site-years, individual blocks contained more than one hybrid/variety. Therefore, it would not be possible to evaluate the interactions of hybrid/variety with planting rate, soil features, and topographical characteristics and their effects on yield.

To address these issues, a finer resolution grid-based system was developed to serve as the experimental units. Square grids were created in QGIS (QGIS Development Team, 2019) with the length and width of the grid cells set equal to the width of the planter, unless a split planter was used to sow multiple hybrids/varieties (Figure 3.4). In that case, the length and width of the grid cells were set to half the width of the planter. Each grid cell was assigned a unique identifier (ID). All data t

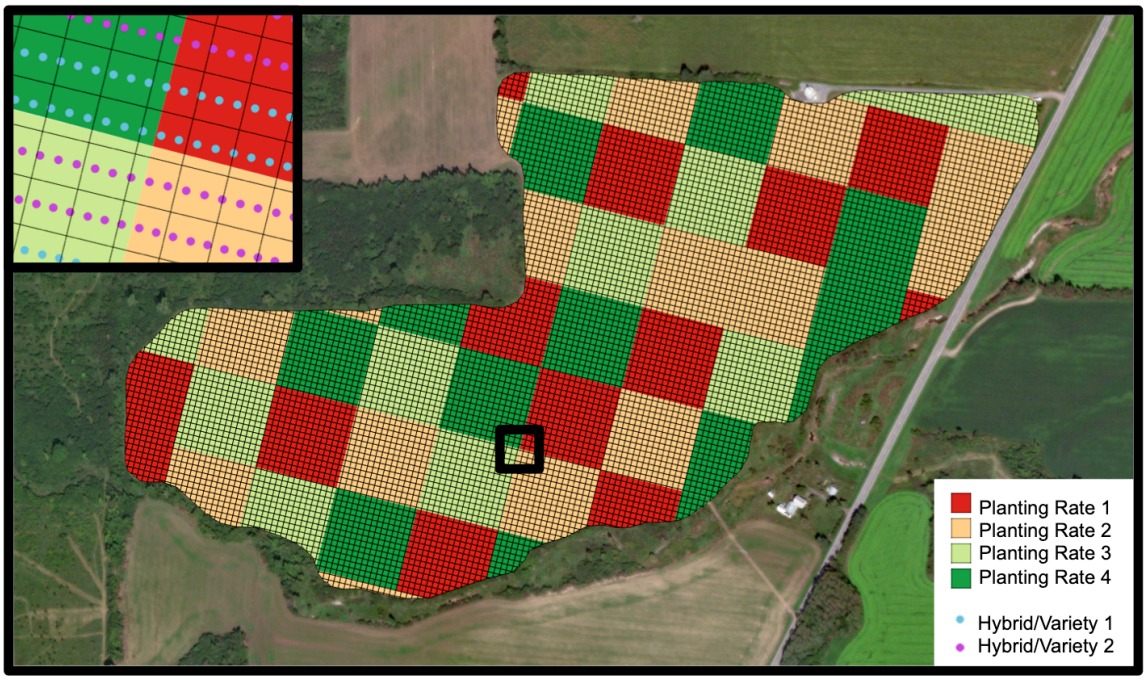


Figure 3.4: Grid of experimental units with respect to the randomized block planting rate design and the alternating hybrids/varieties design

The black box within the field denotes the area represented in the inlay. The length and width of the grid cells in this example are equal to half the width of the split planter such that data from only one hybrid/variety are present within each cell.

ypes were aggregated to the experimental unit grid to resolve the data into a tabular format for analysis.

Data Types and Quality Control

For each site-year, five spatial data types were used to assess the relationship between yield, planting rate, hybrid/variety, topographical characteristics, and soil characteristics. The layers are as follows: 1) harvest, 2) target and as-applied planting, 3) topography, 4) grid soil sampling, and 5) Natural Resources Conservation Service (NRCS) Soil Survey (Figure 3.5).

The harvest point layer recorded information about the levels of dry volume yield harvested throughout the field. In the form of polygons, the target planting rate layer represented the variable rate planting design by which the growers sowed the trials, while the as-applied planting rate point layer contained information about the precise rates applied by the planter as well as where each hybrid/variety was sown. These three layers were used in conjunction with one another to perform various quality control measures to the as-applied and yield data (Figure S3.1A-F). Briefly, the headlands were removed from both the as-applied planting and harvest layers in QGIS. The remaining data points were then assigned ID numbers corresponding to the experimental unit grid cell in which they appeared. The as-applied planting and harvest layers were then analyzed in R (R Core Team, 2019) to remove outlying grid cells based on the mean, variance, and number of data points within each cell. Grid cells were also removed from the analysis if they contained data corresponding to more than one target planting rate or hybrid/variety. Finally, grid cells in which the as-

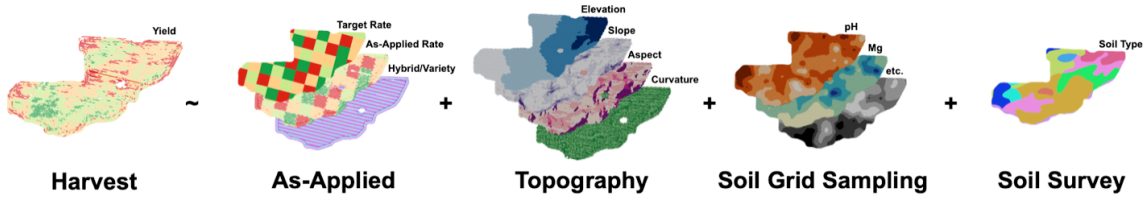


Figure 3.5: The five spatial data types used to model the relationship between yield, planting rate, hybrid/variety, topographical features, and soil characteristics

applied planting rate greatly deviated from the target planting rate were also removed.

Topographical variables for elevation, slope, aspect, and curvature were derived from elevation measurements recorded by the GPS systems aboard the planting machinery (Figure S3.1G). Using the as-applied planting rate layer, elevation means were calculated in QGIS using all of the data points within each experimental grid cell. Slope, aspect, and curvature were then calculated in R using 3×3 neighborhoods of cells following Burrough and McDonell (1998) and Zevenbergen and Thorne (1987).

Integrated Ag Services (Mildford Center, OH) conducted grid soil sampling along a 0.20-ha grid pattern, and samples of the topsoil were analyzed by. Spectrum Analytic (Washington Court House, OH). Point data on up to 12 soil features were provided for each field: aluminum (Al), phosphorus (P), potassium (K), potassium saturation (KSt), calcium (Ca), calcium saturation (CaSt), magnesium (Mg), magnesium saturation (MgSt), pH, buffer pH (BpH), cation exchange capacity (CEC), and soil organic matter (OM). K, Mg, and CEC were recorded for all 27 locations (57 site-years). The remaining variables were unavailable for ≤ 4 locations (≤ 12 site-years) with the exception of Al, which was available for only 17 locations of the 27 locations (37 site-years). The available soil variables for each location are illustrated in Figure 3.6. Using the "gstat" package in R, the soil samples were block kriged to the experimental unit grid (Figure S3.1H). If a semivariogram for a particular variable could not be fit due to insufficient spatial autocorrelation, the values for that variable were set to missing.

The NRCS Soil Survey layer in the form of polygons was obtained for each

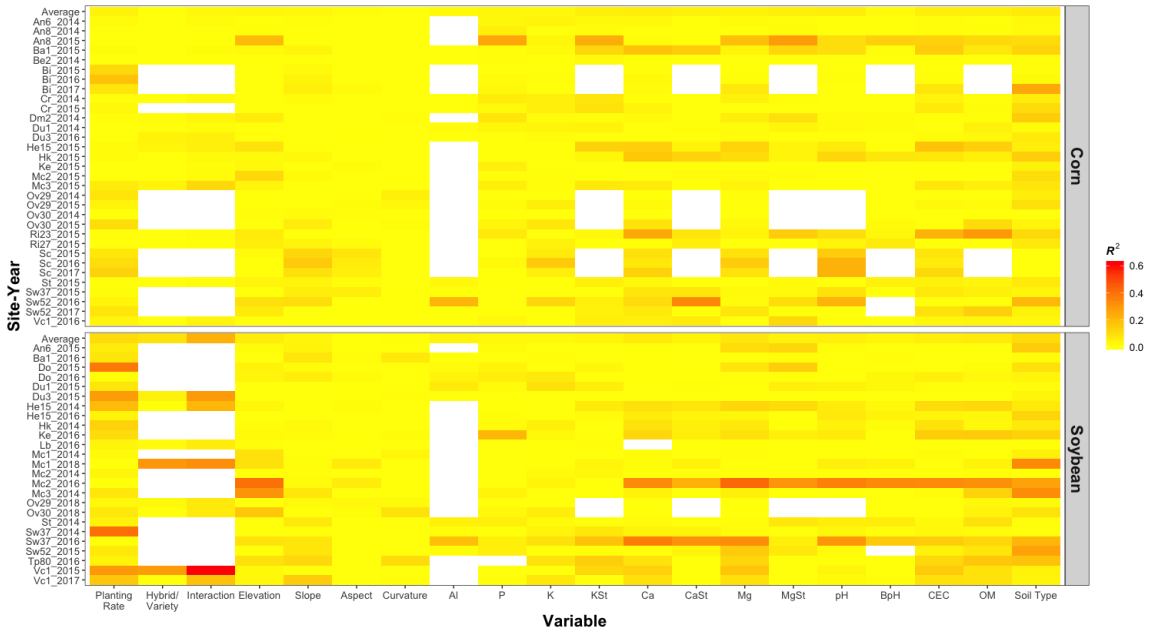


Figure 3.6: R^2 values of yield regressed onto each variable in each site-year with simple linear regression

"Interaction" refers to the R^2 of yield regressed onto planting rate, hybrid/variety, and their interaction. The top row for each crop is the average over all site-years of that crop. White indicates that the variable was not measured or could not be included in the analysis due to low spatial autocorrelation.

field from the SSURGO database (Soil Survey Staff, Natural Resources Conservation Service, USDA, 2015). The soil type attribute was extracted for each experimental grid cell using QGIS. If an experimental unit grid cell contained more than one unique soil type, the soil type for that cell was set to missing (Figure S3.1I)

Final datasets containing variables describing the yield, target and as-applied planting rate, topography, grid soil sampling, and NRCS Soil Survey were merged into a combined data frame in R using the experimental unit grid cell IDs. Excluding yield as the response variable, the dimensions of the final datasets for each site-year used for model fitting were $n \times p$ where n is the number of experimental unit grid cells and p is the number of predictors.

Variable Correlation, ANOVA, and Linear Regression

Pearson's correlations were calculated between all variables. Treatment effects for planting rate and hybrid/variety were estimated using ANOVA. Yield was regressed on each topographical and soil variable to estimate the percent of yield variation explained (R^2). Yield was also regressed on planting rate, hybrid/variety, and their interaction to assess their collective influence on yield in a linear context. All correlation, ANOVA, and linear regression analyses were carried out in base R.

Random Forest Regression

Random forest regression models were fit to assess prediction accuracy, calculate variable importance, and develop optimized planting rate designs. Forests were grown using the "cforest" function of the "party" package for R (Hothorn et al., 2006; Strobl et al., 2007; Strobl et al., 2008). Through simulation, Strobl et al. (2007)

demonstrated that bias can be observed in the variable importance measures when predictors vary in their scale of measurement or, for categorical variables, in their number of levels. To avoid bias, Strobl et al. (2007) proposed "cforest" as an alternative implementation of the commonly used regression functions in the "randomForest" package. The algorithm is based on conditional inference trees and applies subsampling without replacement, which was shown to produce reliable variable importance measures when predictors varied in their scale of measurement or number of categories.

Random forest regression models were fit using each site-year as the training set. The hybrid/variety variable was excluded as a predictor because predictions could not be made across site-years for which different hybrids/varieties were sown. The models may therefore be described as "agnostic" to the hybrid/variety planted. Fitted models were then used to predict yield for each remaining site-year planted with the same crop as the test set. If the soil type for a given experimental unit grid cell in the test set was not observed in the training set, soil type for that grid cell was set to missing. Prediction accuracy was assessed as the Pearson's correlation between the predicted and observed values for yield. Once prediction accuracy had been assessed for all pairs of site-years, the mean predictive ability was calculated for each site-year as the training set considering its prediction accuracies for all other sites-years.

Variable importance measures for each predictor were calculated for each site-year using the "varimp" function in the "party" package for R. Briefly, variable importance is calculated by permuting each predictor variable to determine the difference in prediction accuracy before and after the permutation. Variable

importance measures were scaled to enable comparisons across site-years. To serve as a metric for the similarity between pairs of site-years, Euclidean distances between site-years in terms of their scaled random forest variable importance measures were calculated using the "dist" function in R.

A subset of the full dataset was utilized to evaluate the effect of the *mtry* hyperparameter on model accuracy (data not shown). In summary, prediction models were trained for each site-year using all possible values for *mtry* $\{1 \ldots p\}$. The fitted models were then applied to the remaining site-years in the subset planted with the same crop. Overall, prediction accuracies showed minimal variation among the evaluated *mtry* values. For each fitted model, there was no *mtry* value that consistently provided the highest accuracies across the predicted site-years. Therefore, the default *mtry* value of $p/3$ was used for this study (Breiman, 2001). The *ntree* value was set to 1000.

Optimized Planting Rate Designs

Optimal planting rate designs were developed for each site-year using its own data as the training set to fit the random forest regression model. For each experimental unit grid cell, trained models were used to predict yield at each of the four levels of the evaluated planting rates, given the underlying soil and topographical features of that grid cell. The designs were first developed to optimize planting rates for yield by identifying the planting rate for each grid cell that provided the highest yield prediction. This approach was also extended to consider seed costs and market prices in order to optimize planting rate designs for maximum profits. Planting rates were optimized according to theoretical "favorable," "average," and "unfavorable"

scenarios with respect to the ratio between the cost of seed and the price of the harvest grain (herein referred to as cost:price ratio). For corn, the "favorable," "average," and "unfavorable" cost:price ratios used were 0.5, 1, and 1.5, respectively, while for soybean, the ratios were 0.03, 0.04, and 0.05, respectively. Planting rate designs developed for the same site-year were compared to one another by calculating the proportion of grid cells for which the assigned planting rates were the same.

Results

Grain Yield Summary Statistics

The average grain yields across all site-years were 11,566 kg ha^{-1} for corn and 3,676 kg ha^{-1} for soybean, which were above the average statewide yields of 9,953 and 3,026 kg ha^{-1} for corn and soybean, respectively, (USDA-NASS, 2019) during the 2014-2018 period. Corn yields were lowest and most variable during the 2015 growing season, which was characterized by extremely wet conditions during May and June followed by dry conditions from early July through mid-September. Scouting reports recorded in early July noted multiple fields with standing water in some sections (data not shown). Corn yields were likewise variable in 2016, due to severe drought conditions in June and July in western and central New York. For soybean, 2015 yields were the most variable with respect to their within-site-year standard deviations, while yields were the lowest in 2016.

Effect of Planting Rate and Hybrid/Variety on Yield

Although the ANOVA of planting rate showed a significant difference in yield between at least one pair of planting rate levels (p-value < 0.05) for 31 out of 32 site-years for corn, planting rate on average explained only 4.1 percent of yield variation

(Figure 3.6). The amount of yield variation explained by planting rate ranged from 0.0 to 17.4 percent for corn. For soybean, a significant (p-value < 0.05) treatment effect was observed all 25 site-years, and planting rate explained 10.8 percent of the yield variation for soybean on average. Yield variation explained by planting rate range from 0.6 to 40.8 percent for soybean, and values ≥ 10 percent were observed in 8 of the 25 site-years.

Significant treatment effects (p-value < 0.05) were observed for hybrid/variety in 13 out of the 19 site-years for which multiple hybrids of corn were sown and for 5 out of 9 site-years for which multiple varieties of soybean were sown. The average percent of yield variation explained by hybrid/variety was 1.0 percent for corn and 8.5 percent for soybean. The percent of yield variation explained by hybrid/variety ranged from 0.0 to 4.1 percent for corn and from 0.0 to 30.1 percent for soybean. Notably, hybrid/variety explained over 25 percent of variation in soybean yields in 2 site-years: Vc1_2015 and Mc1_2018. In Vc1_2015, yields for variety P92Y51 exceeded those of P22T41 by 771 kg ha^{-1}, on average. In Mc1_2018, yields for soybean variety SG1776 exceeded those of AG3334 by an average of 395 kg ha^{-1}.

The interaction between planting rate and hybrid/variety was significant (p-value < 0.05) in 18 out of the 19 site-years for which multiple hybrids of corn were sown and for all 9 site-years for which multiple varieties of soybean were sown. The percentages of yield variation explained by planting rate, hybrid/variety, and their interaction had means and ranges of 2.8 and 0.2 to 11.5 percent for corn and 22.8 and 6.2 to 61.9 for soybean, respectively. Among corn site-years, Mc3_2015 showed the highest percent yield variation explained by planting rate, hybrid/variety, and their

interaction, at a level of 11.5 percent. The hybrids grown for Mc3_2015 were P0506AM and P0533AM1. Both showed a positive linear response to planting rate, with yields for P0506AM exceeding those of P0533AM1 by an average of 523 kg ha^{-1}. For soybean, planting rate, hybrid/variety, and their interaction explained 61.9 percent of the variation in yield for Vc1_2015. Both varieties P22T41 and P92Y51 showed positive linear responses to planting rate.

Topographical and Soil Summary Statistics

The 27 field locations tested varied greatly in their topographical and soil feature profiles. The lowest and highest elevations recorded were 105 and 368 masl, and the difference between lowest and highest points in a field on average was 16 m. The largest gradient within a single field with respect to elevation was observed at Ke, which ranged from 116 to 164 masl for a difference of 48 m. The United States Environmental Protection Agency classifies crop production on slopes ≥ 9 percent, or 5.14°, as "agricultural land cover on steep slopes." Of the 27 field locations, 19 contained areas of sloped terrain at grades of 5.14° or above. The field location Be2 had a sloped terrain feature with a grade of 20°.

Soils in New York State are generally acidic. The soil grid sampling data showed that the pH for the 27 tested field locations ranged from 5.8 to 7.2. High levels of aluminum were likewise observed, ranging from 657 to 767 ppm. CEC levels were moderate, ranging from 6.2 to 16.5 meq/100 g. Soil OM was moderate to low, averaging 2.3 percent. According to the NRCS Soil Survey, 48 unique soil types were present across the 27 locations. Of the 48 types, 13 were observed in one field only. The average field contained 6 soil types, while Du3 contained the most at 13 types.

"Cazenovia" was the most frequently observed soil type, appearing in 10 of the field locations.

The upper triangle of Figure 3.7 shows the average Pearson's correlation between each pair of topographical and soil characteristics for the 27 locations. The lower triangle corresponds to the standard deviation of the Pearson's correlations. High magnitude means and low standard deviations for a pair of characteristics are desirable for a prediction context because this would indicate that the two variables are not only strongly related, but their relationship is consistent across years and sites. As expected, pH and Al were observed to have a strong negative correlation, while CEC and OM were positively correlated. The relationship between OM and elevation was observed to be slightly negative, on average, although highly variable across the 27 locations.

Fig 3.6 shows the percent of yield variation explained by each of the topographical and soil features across the 57 site-years. When averaging over the 32 site-years for corn and 27 site-years for soybean, no variables with the exception of soil type for soybean explained more than 10 percent of the variation in yield, on average, suggesting that no topographical or soil variables were consistently predictive of grain yield across site-years. Soil type was observed to have the highest R^2 values among the topographical and soil variables for both corn and soybean, explaining 6.8 and 10.4 percent of the variation in corn and soybean yields, respectively.

Random Forest Regressions

A random forest regression model was fit to predict yield using each site-year as the training set. The average RMSEs were 1,465 kg ha^{-1} for corn and 162 kg ha^{-1} for

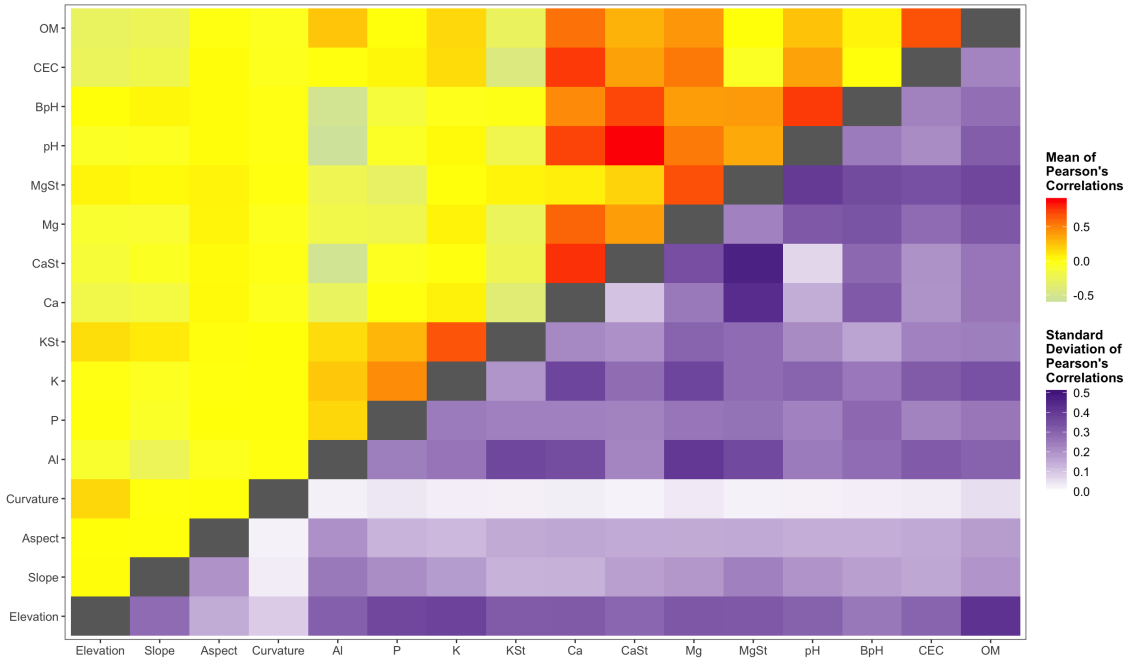


Figure 3.7: Means and standard deviations of the Pearson's correlations between each pair of variables across the 27 field locations

soybean. The R^2 values ranged from 0.16 to 0.78, with an average of 0.47, for corn and from 0.17 to 0.72, with an average of 0.50, for soybean. With respect to the top four site-years for corn in terms of random forest regression R^2 values ($R^2 \geq 0.7$), planting rate explained only 0, 1, 4, and 7 percent of the variation in yield when tested in simple linear regression. Conversely, for soybean, in two of the top four site-years with respect to random forest regression R^2 values ($R^2 \geq 0.6$), planting rate explained 30 and 38 percent of the variation in yield in linear regression. For the two other site-years, planting rate explained no variation in yield in linear regression.

The scaled variable importance measures for each predictor across the 32 corn and 25 soybean site-years are illustrated in Figure 3.8. Planting rate consistently had higher importance compared to other variables for both crops but particularly for soybean. Elevation was likewise important for both crops in most site-years. Organic matter appeared to be more consistently important for soybean than for corn. The soil nutrient variables were more frequently of greater importance for corn than for soybean. Most notably, many site-years had one or more variables with a high level of importance such that there was little consistency observed across site-years in terms of variable importance rankings.

Yield Predictions Across Environments

When random forest regression was applied to predict yield across site-years, the prediction accuracies varied greatly based on the pair of site-years used to train and test (Figure 3.9). This is evident in the mean prediction accuracy for each site-year, which average 0.03 for corn and 0.08 for soybean. However, when considering pairs separately, each site-year predicted at least one other site-year at a level of ≥ 0.20

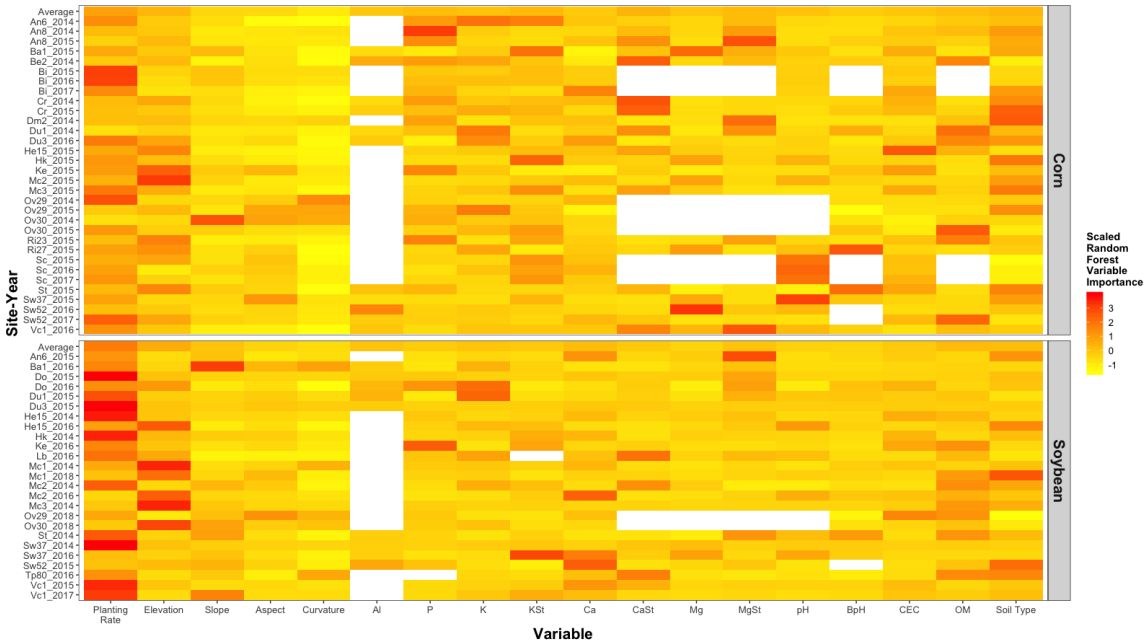


Figure 3.8: Scaled random forest regression variable importance measures for each variable in each site-year

The top row for each crop is the average variable importance over all site-years of that crop, White indicates that the variable was not measured or could not be included in the analysis due to low spatial autocorrelation.

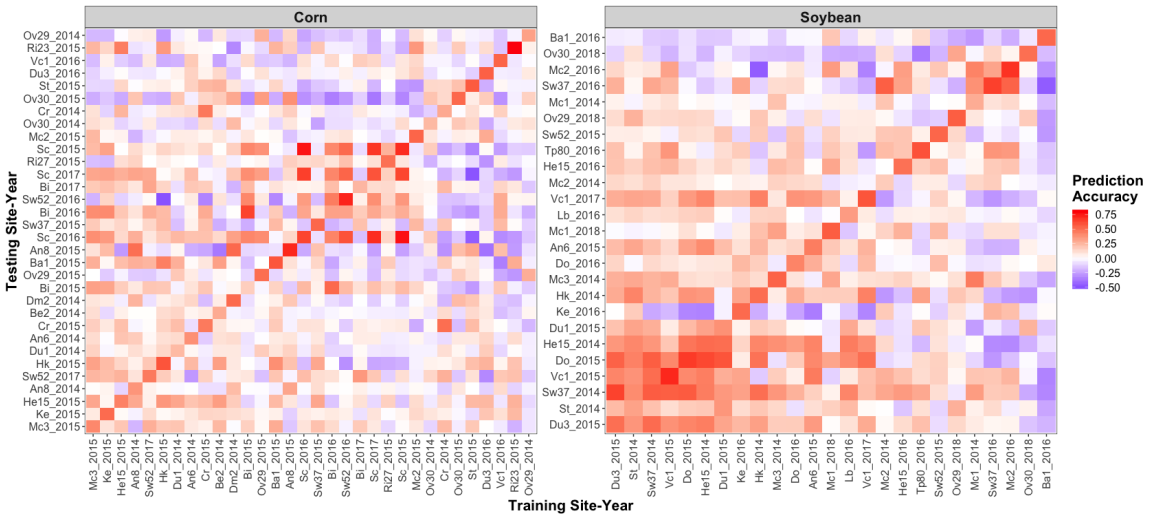


Figure 3.9: Across site-year prediction accuracy of random forest regression models

Site-years are ordered with respect to their mean prediction accuracy across all other site-years. The diagonal corresponds to the R^2 of the random forest regression training model for each site-year.

(with site-year Ba1_2016 removed due to very low across-site-year prediction accuracy). The site-years with the highest mean prediction accuracies across all other site-years were Mc3_2015 for corn at a level of 0.11 and Du3_2015 for soybean at a level of 0.20. The site-year pairs with the highest accuracies were generally those for the same field location tested during different years. For example, site-year Bi_2015 had an average prediction accuracy of 0.05 across all site-years but predicted Bi_2016 with an accuracy of 0.64. For field locations that were tested for the same crop under two different years, the target planting rate design was re-randomized from one year to the next with the exception of Sc_2015 and Sc_2016. Those two site-years were sown using the same planting rate randomization in both 2015 and 2016. Notably, Sc_2015 and Sc_2016 predicted each other with accuracies of 0.76 and 0.75, respectively. Slightly lower accuracies of 0.63 and 0.65 were observed, respectively, when predicting Sc_2017, which was planted under a re-randomized design. The prediction accuracies for each pair of site-years were correlated with the Euclidian distances between the variable importance measures for each pair of site-years at a level of -0.23 for corn and -0.48 for soybean, suggesting that site-year pairs with similar variable importance profiles tend to predict one another at a higher level of accuracy.

Planting Rate Optimization

Optimized planting rate designs were developed from the random forest regression prediction models to maximize yields as well as profits based on varying cost:price ratios. Figure 3.10 illustrates these designs for an example site-year, and Figure 3.11 shows the proportions of each optimized planting rate design assigned to

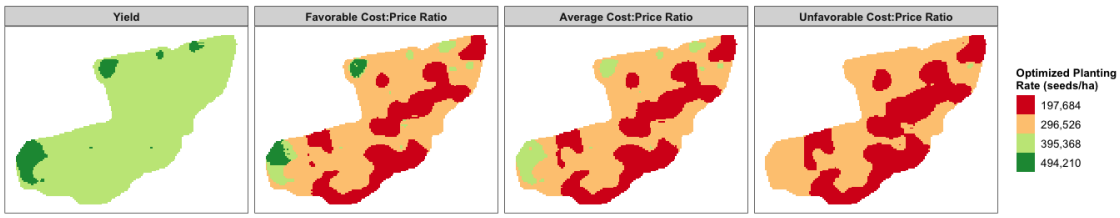


Figure 3.10: Example of the planting rate designs optimized yield and for profit at "favorable", "average", and "unfavorable" cost:price ratios

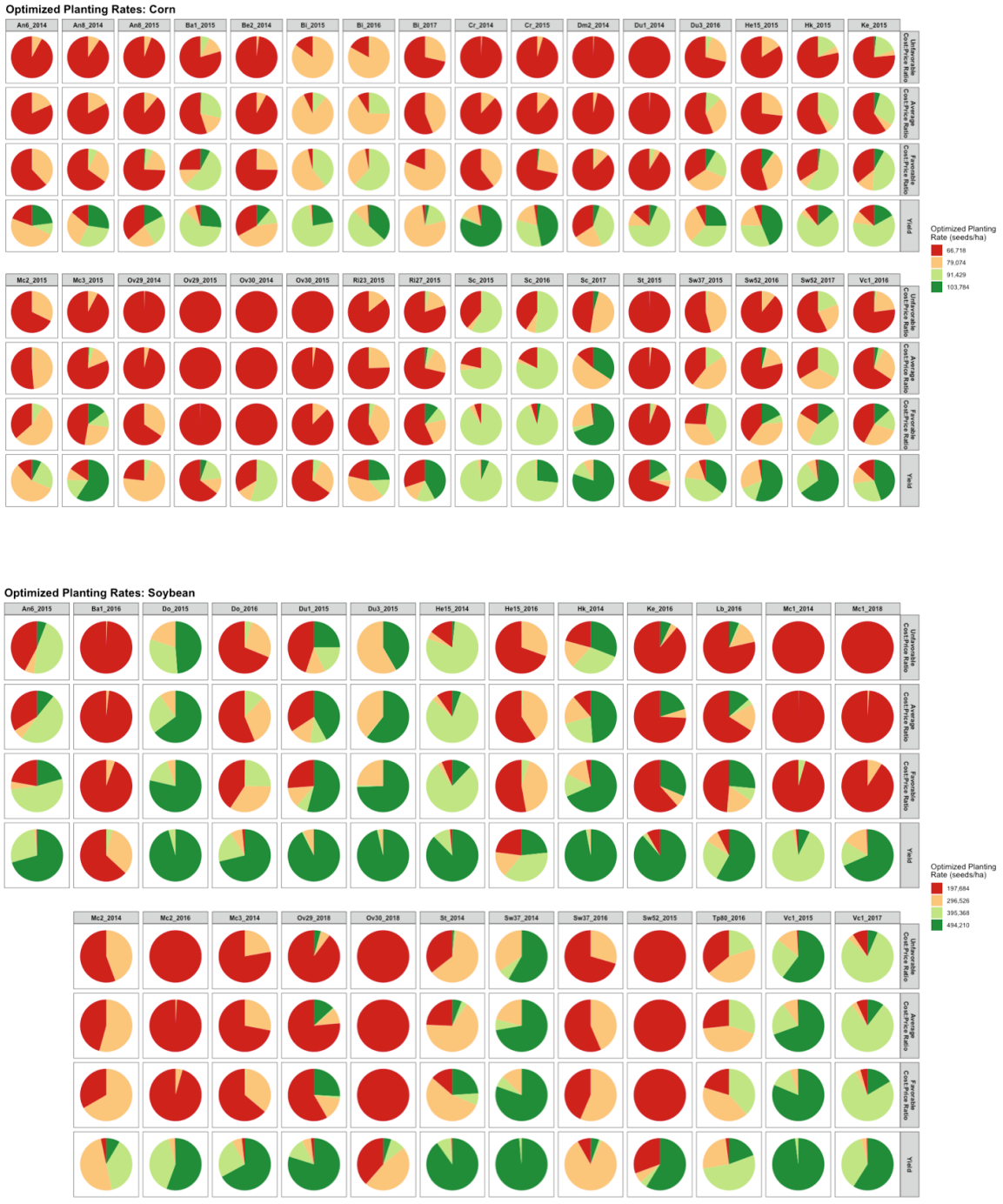


Figure 3.11: Proportions of each optimized planting rate design assigned to each of the planting rate levels

each of the planting rate levels. In general, optimizing for yield was the most liberal approach with respect to assigning planting rate levels, favoring higher rates compared to optimizations for profit. For corn, the four planting rates in increasing order were assigned to, on average, 18.6, 21.8, 33.3, and 28.2 percent of the field. For soybean, optimizing for yield overwhelming favored the highest rate, which was assigned to 60.1 percent of the field, on average. The remaining three rates in increasing order were assigned to, on average, 9.7, 13.8, and 18.9 percent of the field.

A ratio of the seed cost to the crop market price was developed to assess planting rate optimizations with respect to overall profits. In general, optimizing for the cost:price ratio increased the frequency of lower rates compared to the optimizations for yield (Figure 3.10). For some site-years, the contrast between the yield and cost:price ratio optimizations were stark. For example, for Cr_2014, optimizing for yield overwhelming favored the highest rate, which was assigned to 80.8 percent of the field. However, when optimizing for the cost:price ratio, the lowest rate was greatly favored, having been assigned to 60.3, 88.3, and 99.2 percent of the field for the "favorable," "average," and "unfavorable" cost:price ratio scenarios, respectively. For other site-years, the contrasts between yield and cost:price ratio optimizations were less pronounced. For example, for Ri27_2015, the four rates in increasing order were assigned to 30.2, 13.2, 14.3, and 42.3 percent of the field when optimizing for yield. The distribution of rates in increasing order for the "favorable" cost:price ratio scenario were 56.8, 22.1, 10.4, and 10.7 percent.

Overall, the optimized designs for corn were assigned primarily variable rates when optimizing for yield or for profit at a favor cost:price ratio. When the ratio was

"average" or "unfavorable", most designs were predominantly allocated fixed rates. For soybean, the highest rate was assigned to most of the field when optimizing for yield in many site-years. When optimizing for profit, however, the results were less consistent, with some site-years assigned primarily variable rates while others were largely allocated fixed rates.

As expected, the optimizations for yield shared the most similarity with the optimizations for the "favorable" cost:price ratio. For corn, 53 percent of the field, on average, was assigned to the same rate by both optimization schemes. For soybean, an average of 44 percent of the field was assigned to the same rate by both schemes. The similarity between the yield and cost:price ratio optimizations was diminished for the "average" and "unfavorable" scenarios. Notably, for the "unfavorable" scenarios, 28.0 and 27.5 percent of the field on average was assigned to the same rate as the yield optimization for corn and soybean, respectively. The expectation for the similarity between optimizations when rates are assigned at random is 25 percent.

The cost:price ratio optimizations were more similar to one another. For corn, 76.7, 86.5, and 64.9 percent of the field, on average, was assigned to the same rate for the "favorable" and "average," "unfavorable" and "average," and "favorable" and "unfavorable" optimization pairs, respectively. For soybean, 87.9, 89.1, and 77.7 percent of the field, on average, was assigned to the same rate for the respective optimization pairs.

For corn, 7 of the field locations were evaluated for the same crop in more than one year. Among these, 2 were evaluated over 3 years while the remaining 5 were evaluated over 2 years. For soybean, 6 field locations were evaluated over 2 years.

When comparing yield optimizations across years, the proportion of the field assigned to the same rate was 31.5 percent on average for corn, with a range of 12.7 to 44.3 percent. For soybean, the average was 33.2 percent with a range of 5.3 to 74.3 percent. For the cost:price ratio, the average proportions of the field assigned to the same rates across years were 46.8, 56.5, and 64.7 percent for the "favorable", "average", and "unfavorable" scenarios, respectively, for corn. For soybean, the averages were 26.8, 30.9, and 33.1 percent, respectively.

Discussion

Previous efforts to evaluate the potential for variable rate planting in corn and soybean have largely focused on the relationship between planting rate and yield in relatively few site-years. Consequently, the results have been mixed, with some studies reporting potential for profitability (Shanahan et al., 2004; Hörbe et al., 2013; Trevisan et al., 2019), while others have ascribed inconsistencies in results to environmental and weather variation across sites and years (Cox and Cherney, 2012; Reeves and Cox, 2013; Smidt et al., 2016; Licht et al., 2017). To our knowledge, this study is the first to evaluate the potential for variable rate planting within a broader inference space, encompassing a large number of diverse locations sampled across years. Consequently, the conclusions developed here may begin to describe the distribution of possible site-year conditions as opposed to the behavior of variable rate in specific site-years.

The first objective was to evaluate the interactions of planting rate with topographical and soil features and as well as their effects on yield in corn and soybean. In the context of linear modeling, planting rate explained relatively low

amounts of the yield variation, particularly for corn. Low *F*-ratio test statistics observed for the planting rate treatment effects implies that there is substantial variation in grain yield that can be attributable to other sources of variability, such as topographic and soil characteristics, which may or may not be interacting with planting rate to drive differences in yield. In roughly half of the site-years for corn, a topographical or soil variable explained over 10 percent of the variation in grain yield in the linear context.

Though in most cases there was no individual planting rate that would have given a substantial yield advantage over others in a linear context for corn, planting rate had the highest average importance among all of the predictor variables in random forest regression for both corn and soybean. This contrast in results may suggest that there are important complex and/or nonlinear interactions between planting rate and other variables that are not being captured through linear modeling. This result is consistent to a similar study that found artificial neural networks, which like random forest regression are capable of modeling nonlinear interactions, to be superior to linear modeling approaches when predicting corn yields with soil and landscape features (Miao et al., 2006). Random forest regression is among a suite of established and emerging techniques in machine learning that are becoming more widely used in agricultural research to model complex systems (Henderson et al., 2005; Häring et al., 2012; Xiong et al., 2014; Chlingaryan et al., 2018; Liakos et al., 2018). As more environmental, climate, management, and economic data of varying types and scales become available to growers through innovations in precision agriculture, it will be important for the agronomy community to acknowledge the limitations of linear

modeling when working with large numbers of complexly related variables and to begin to leverage the advances made in machine learning-based approaches.

The proposed random forest regression-based method for optimizing variable rate plating designs is flexible to allow for the maximization of either yield or profit with respect to user-defined seed costs and crop market prices. Marked differences were observed between corn and soybean with respect to their planting rate optimizations. For corn, the designs optimized for yield were largely variable with all four planting rate levels assigned to some proportion of the field in 31 of the 32 site-years. However, when optimizing for the cost:price ratio, the designs were distinctly less variable. A single rate was assigned to 90 percent or more of the field for 6, 8, and 14 site-years according to the "favorable," "average," and "unfavorable" cost:price ratios, etc. Conversely, for soybean, single rate was assigned to at least 70 percent of the field for over half of the site-years when optimizing for yield. However, when optimizing for cost:price ratio, the designs were more variable for roughly half of the site-years, while the remaining site-years' designs were more similar to a fixed rate.

These results provide important insights into the relationship between planting rates, yields, and profits that may serve to inform management decisions. The differences in variability between the designs optimized for yield and profit for corn suggest that the marginal gains in yield from planting at higher rates may only be profitable when the ratio of the cost of seed and the price of corn is favorable to growers. As this ratio becomes less favorable, the optimized designs trend towards fixed or near-fixed rates consisting of the lowest planting rates. Growers should therefore factor the current seed costs and commodity prices into the decision to plant

corn at a variable rate.

For soybean, however, fixed or near-fixed rates appeared to provide the maximum predicted yields for most site-years. Planting rate explained a much larger percentage of the variation in soybean yield in the linear modeling context as compared to corn, and it likewise exhibited a high level of variable importance in random forest regression. In terms of maximizing yields, these results suggest that it may be more beneficial to sow soybean at a fixed rate, though identifying the appropriate rate given the selected variety and field conditions appears to be important, as the dominant assigned rate was not necessarily the same across all site-years. For maximizing profits, the extent of variability in the optimized designs was not consistent across site-years. In addition, for field locations that were assessed in more than one year, the cost:price ratio-based designs were less consistent across years for soybean than compared to corn. This inconsistency could be due to variations in environmental conditions across site-years (Wells, 1993), differential sensitivities to planting rate among varieties sown (Agudamu et a., 2016), interactions between planting rate and row spacing (Cox and Cherney, 2011), etc.

Variations in optimal planting rate designs were likewise observed for fields that were evaluated under the same crop in multiple years. The proportion of the field assigned to the same planting rate when using each year's data to build the optimized designs was low to moderate. This was particularly true for soybean: optimizations for the same field developed from different years' data shared roughly 30 percent identity in assigned rates, which is comparable to the expectation of similarity under when rates are assigned at random. Further long-term testing is needed to more accurately

assess the extent to which year-to-year variation impacts optimized designs for both corn and soybean. Additionally, while this study has aimed to broadly assess the potential for the profitability of variable rate planting in corn and soybean in a data-driven framework, empirical studies are greatly needed to validate the methods proposed herein. The New York Corn and Soybean Growers Association is currently conducting on-farm validation experiments of the random forest regression-based planting rate optimization approach.

Lastly, the expense of implementing the proposed approach is not trivial. Grid soil sampling was conducted at each field location, and research associates trained in data science were needed to process and analyze the data. In addition, an added bottleneck of this approach, in its current form, is that the optimized planting rate designs were developed from random forest regression models trained on the data observed in the field location which was to be predicted. This requires growers to plant and evaluated the variable rate randomized block design for at least one year before optimized rates can be determined.

While local testing is likely to always provide the most accurate optimized rates, the ability to extrapolate optimization models onto untested fields would be of great interest. Overall, using the fitted random forest regression models to predict across site-years yielded low to moderate prediction accuracies. The site-years with the highest mean prediction accuracies across all other site-years were Mc3_2015 and Du3_2015 with accuracies of 0.11 and 0.20, respectively. Low accuracies are perhaps expected given the variation across site-years in the direction and magnitude of the relationships between predictor variables. However, there may be alternative

approaches for building the most useful training sets for a given untested field. For example, there was a significant negative correlation between the across site-year yield prediction accuracies and the Euclidean distances between site-years in terms of random forest regression variable importance. Provided yield, topographical, and soil data are available for an untested field, random forest regression models predicting yield could be fit, and the resulting variable importance measures could be utilized to identify the most similar site-year evaluated under the randomized block design to use as a training set. Additionally, data from multiple site-years could be combined in an optimal fashion to maximize prediction accuracy. Heslot et al. (2013) provides an innovative approach for this type of training set optimization in the context of plant breeding that identifies and removes less predictive site-years from the complete set of site-years used to train a combined model.

Conclusions

In this study, we have proposed a random forest regression-based approach to predict yields given high resolution topographical, soil, and variable rate planting rate data. We have likewise developed a method for optimizing planting rate designs to maximize fields or profits according to user-defined seed costs and crop market prices. Prediction models were trained using on-farm corn and soybean data collected from 57 site-years evaluated in New York State between 2014 and 2018. While planting rate explained relatively low amounts of yield variation in the linear context, particularly for corn, planting rate was ranked most highly on average in terms of random forest regression variable importance, which is capable of accounting for non-linear and complex interactions between variables. This demonstrates a potential

limitation of linear modeling within in the context of highly dimensional agricultural datasets. The optimized planting rate designs suggested that variable rate may be useful for maximizing yields or profits for corn when the seed cost to crop market price ratio is favorable to growers. As the ratio becomes less favorable, it may be more advantageous to plant at lower fixed rates. For soybean, the optimized planting designs appeared to favor fixed or near-fixed rates for maximizing yields, though the predominant rates were not necessarily consistent across site-years. Results were more variable when optimizing planting rate designs to maximize profits, indicating that the marginal benefit of increased planting rates may be influenced by other factors such as environmental conditions, differential sensitivity of varieties to planting rate, interactions between planting rate and row spacing, etc. Further long-term studies will be necessary to evaluate the stability of optimized designs across years. Likewise, empirical studies are needed to evaluate the profitability of the proposed method; on-farm validation experiments are currently ongoing. Lastly, yield prediction across site-years appears to be relatively poor, perhaps due to the wide variability of the magnitude and direction of the relationships between variables. Local testing is likely needed to provide the most accurate planting rate optimizations; however, further research is needed to evaluate training set optimization.

REFERENCES

Agudamu, T. Yoshihira, and T. Shiraiwa, 2016 Branch development responses to planting density and yield stability in soybean cultivars. Plant Prod. Sci. 19: 331-339.

Archontoulis, S.V., and F.E. Miguez, 2014 Nonlinear regression models and applications in agricultural research. Agron. J. 107: 786-798.

Barnhisel, R.I., M.J. Bitzer, J.H. Grover, and S.A. Shearer, 1996 Agronomic benefits of varying corn seed populations: a Central Kentucky study. In: P.C. Robert, R.H. Rust, and W.E. Larson (Eds.), *Precision Agriculture*, ASA, CSSA, SSSA, Madison, WI. p. 957-965.

Breiman, L., 2001 Random forests. Mach. Learn. 45: 5-32.

Brubacker, S.C., A.J. Jones, D.T. Lewis, and K. Frank, 1993 Soil properties associated with landscape position. Soil. Sci. Soc. Am. J. 57: 235-239.

Bullock, D.G., D.S. Bullock, E.D.Nafziger, T.A. Doerge, S.T. Paskiewicz, P.R. Carter, and T.A. Peterson, 1998 Does variable rate seeding of corn pay? Agron. J. 90: 830-836.

Burrough, P.A., and R.A. McDonnell, 1998 *Principles of Geographic Information Systems.* New York: Oxford University Press.

Changere, A., and R. Lal, 1997 Slope position and erosional effects on soil properties and corn production on a Miamian soil in central Ohio. J. Sustain. Agr. 11: 5-21.

Chlingaryan, A., S. Sukkarieh, and B. Whelan, 2018 Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture.

Comput. Electron. Agr. 151: 61-69.

Corassa, G.M, T.J.C. Amado, M.L. Strieder, R. Schwalbert, J.L.F. Pires, P.R. Carter, and I.A. Ciampitti, 2018 Optimum soybean seeding rates by yield environment in Southern Brazil. Agron. J. 110: 2430-2438.

Cox, M.S., P.D. Gerard, M.C. Wardlaw, and M.J. Abshire, 2003 Variability of selected soil properties and their relationships with soybean yield. Soil Sci. Soc. Am. J. 67: 1296-1302.

Cox, W.J., and J.H. Cherney, 2011 Growth and yield responses of soybean to row spacing and seeding rate. Agron. J. 103: 123-128.

Cox, W.J., and J.H. Cherney, 2012 Lack of hybrid, seeding, and nitrogen rate interactions for corn growth and yield. Agron. J. 104: 945-952.

Frogbrook, Z.L., M.A. Oliver, M. Salahi, and R.H. Ellis., 2002 Exploring the spatial relations between cereal yields and soil chemical properties and the implications for sampling. Soil Use Manage. 18: 1-9.

Graham, M.H., 2003 Confronting multicollinearity in ecological multiple regression. Ecology 84: 2809-2815.

Häring, T., E. Dietz, S. Osenstetter, T. Koschitzki, and B. Schröder, 2012 Spatial disaggregation of complex soil map units: a decision-tree based approach in Bavarian forest soils. Geoderma 185-186: 37-47.

Henderson, B.L., E.N. Bui, C.J. Moran, and D.A.P. Simon, 2005 Australia-wide predictions of soil properties using decision trees. Geoderma 124: 383-398.

Heslot, N., J.L. Jannink, and M.E. Sorrells, 2013 Using genomic prediction to characterize environments and optimize prediction accuracy in applied

breeding data. Crop Sci. 53: 921-933.

Hörbe, T.A.N., T.J.C. Amado, A.O. Ferreira, and P.J. Alba, 2013 Optimization of corn plant population according to management zones in Southern Brazil. Precis. Agric. 14: 450-465.

Hurley, T.M., K. Oishi, and G.L. Malzer, 2005 Estimating the potential value of variable rate nitrogen applications: a comparison of spatial econometric and geostatistical models. J. Agr. Resour. Econ. 30: 231-249.

Katerji, N., J.W. vann Hoorn, A. Hamdy, F. Karam, and M. Mastrorilli, 1995 Effect of salinity on water stress, growth, and yield of maize and sunflower. Agric. Water Manag. 30: 237-249.

Kitchen, N.R., S.T. Drummond, E.D. Lund, K.A. Sudduth, and G.W. Buchleiter, 2003 Soil electrical conductivity and topography related to yield for three contrasting soil-crop systems. Agron. J. 95: 482-495.

Kravchenko, A.N., and D.G. Bullock, 2000 Correlation of corn and soybean grain yield with topography and soil properties. Agron. J. 92: 75-83.

Lambert, D., J. Lowenberg-DeBoer, and G.L. Malzer, 2006 Economic analysis of spatial-temporal patterns in corn and soybean response to nitrogen and phosphorus. Agron. J. 98: 43-54.

Licht, M.A., A.W. Lenssen, and R.W. Elmore, 2017 Corn (*Zea mays* L.) seeding rate optimization in Iowa, USA. Precis. Agric. 18: 452-469.

Ließ, M., B. Glaser, and B. Huwe, 2012 Uncertainty in the spatial prediction of soil texture: comparison of regression tree and Random Forest models. Geoderma 170: 70-79.

Liakos, K.G., P. Busato, D. Moshou, S. Pearson, and D. Bochtis, 2018 Maching learning in agriculture: a review. Sensors 18: 2674.

Lowenberg-DeBoer, J., 1999 Economics of variable rate planting for corn. In: P.C. Robert, R.H. Rust, and W.E. Larson (Eds.), *Precision Agriculture*, ASA, CSSA, SSSA, Madison, WI. p. 1643-1651.

Miao, Y., D.J. Mulla, and P.C. Robert, 2006 Identifying important factors influencing corn yield and grain quality variability using artificial neural networks. Precis. Agric. 7: 117-135.

Miller, M.P., M.J. Singer, and D.R. Nielsen, 1988 Spatial variability of wheat yield and soil properties on complex hills. Soil. Sci. Soc. Am. J. 52: 1133-1141.

Moore, G.A., and J.P. Tyndale-Biscoe., 1999 Estimation of the importance of spatially variable nitrogen application and soil moisture holding capacity to wheat production. Precis. Agric. 1: 27-38.

Moore, I.D., P.E. Gessler, G.A. Nielsen, and G.A. Peterson, 1993 Soil attribute prediction using terrain analysis. Soil Sci. Soc. Am. J. 57: 443-452.

Mulla, D.J., and J.S. Schepers., 1997 Key processes and properties for site-specific soil and crop management. In: F.J. Pierce, E.J. Sadler (Eds.), *The State of Site-Specific Management for Agriculture*, ASA, CSSA, SSSA, Madison, WI. p. 1-18.

Mutanga, O., E. Adam, and M.A. Cho, 2012 High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm. Int. J. Appl. Earth Obs. Geoinf. 18: 399-406.

Plant, R.E., 2001 Site-specific management: the application of information technology

to crop production. Comput. Electron. Agr. 30: 9-29.

Reeves, G.W., and W.J. Cox, 2013 Inconsistent responses of corn to seeding rates in field-scale studies. Agron. J. 105: 693-704.

Schlenker, W., and M.J. Roberts, 2006 Nonlinear effects of weather on corn yields. Appl. Econ. Perspect. P. 28: 391-398.

Shanahan, J.F., T.A. Doerge, J.J. Johnson, and M.F. Vigil, 2004 Feasibility of site-specific management of corn hybrids and plant densities in the Great Plains. Precis. Agric. 5: 207-225.

Smidt, E.R., S.P. Conley, J. Zhu, and F.J. Arriaga, 2016 Identifying field attributes that predict soybean yield using random forest analysis. Agron. J. 108: 637-646.

Tan, Z.X., R. Lal, N.E. Smeck, and F.G. Calhoun, 2004 Relationships between surface soil organic carbon pool and site variables. Geoderma 121: 187-195.

Tilman, D., K.G. Cassman, P.A. Matson, R. Naylor, and S. Polasky, 2002 Agricultural sustainability and intensive production practices. Nature 418: 671-677.

Trevisan, R.G., D.S. Bullock, and N.F. Martin, 2019 Site-specific treatment responses in on-farm precision experimentation. Preprints 2019020007.

USDA-NASS, 2014-2018 corn and soybean yield statistics for New York. https://quickstats.nass.usda.gov/. Accessed July 2019.

USDA-NASS, 2017 Census of Agriculture, https://www.nass.usda.gov/Publications/AgCensus/2017/. Accessed July 2019.

van Es, H.M., J.D. Woodard, M. Glos, L.V. Chiu, T. Dutta, and A. Ristow, 2016 Digital agriculture in New York State: report and recommendations. Cornell

University, Ithaca, NY.

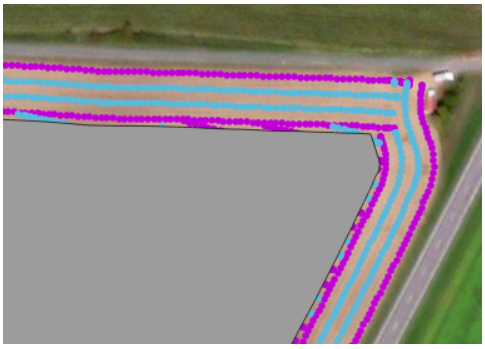
Wells, R., 1993 Dynamics of soybean growth in variable planting patterns. Agron. J. 85: 44-48.

Xiong, X., S. Grunwald, D.B. Myers, J. Kim, W.G. Harris, and N.B. Comerford, 2014 Holistic environmental soil-landscape modeling of soil organic carbon. Environ. Modell. Softw. 57: 202-215.

Zevenbergen, L.W., and C.R. Thorn, 1987 Quantitative analysis of land surface topography, Earth Surf. Proc. Land. 12: 47-56.

Figure S3.1: The pipeline for aggregating the target planting rate, as-applied planting rate, yield, soil survey, and grid soil sampling to a common grid and performing quality control

A Removing the headlands from the as-applied planting rate and yield data layers

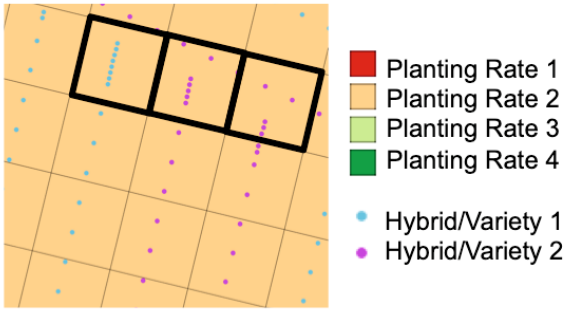


A buffer was created to exclude the headland areas using the “Buffer” function in QGIS.

The headland areas were removed from the as-applied planting rate and yield layers using the “Clip” function in QGIS.

Figure S3.1 (Continued)

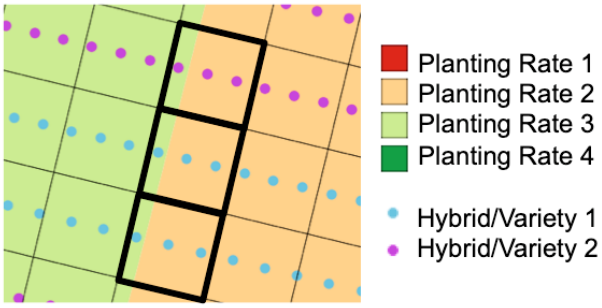
B Removing grid cells based on number of as-applied planting rate or yield data points per cell



Along the edges of the field in particular, the planter and combine often record many data points in rapid succession. To reduce the influence of these data on the analysis, each experimental grid cell was assigned a unique ID. The as-applied planting rate and yield data points within each cell were assigned the respective experimental grid cell ID in QGIS. The number of data points within each grid cell was calculated in R. Significant outliers were identified using Studentized residuals (*p*-value < 0.05) with the lm() and rstudent() functions. Experimental grid cells identified as outliers with respect to the as-planted or yield data layers were removed from the analysis. In the illustrated example, the cells outlined in black boxes would have been identified as outliers and removed from the analysis.

Figure S3.1 (Continued)

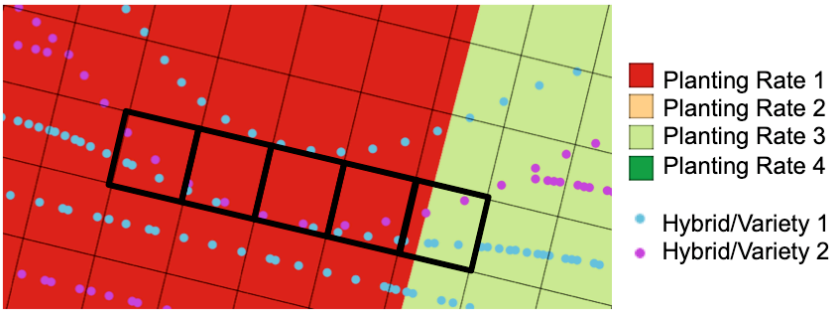
C Removing grid cells that contain data from more than one target planting rate



Because the 0.81 ha target planting rate blocks were not always proportional in size to the experimental unit grid cells, it was not possible to precisely align the experimental unit grid cells with the target planting rate blocks. This resulted in some grid cells containing data corresponding to more than one target planting rate. Using the unique IDs matching data points to grid cells, the target planting rate for each data point within each grid cell was identified in R. Cells that contained data points corresponding to more than one target planting rate were removed from the analysis. In the illustrated example, the cells outlined in black boxes would have been identified and removed from the analysis.

Figure S3.1 (Continued)

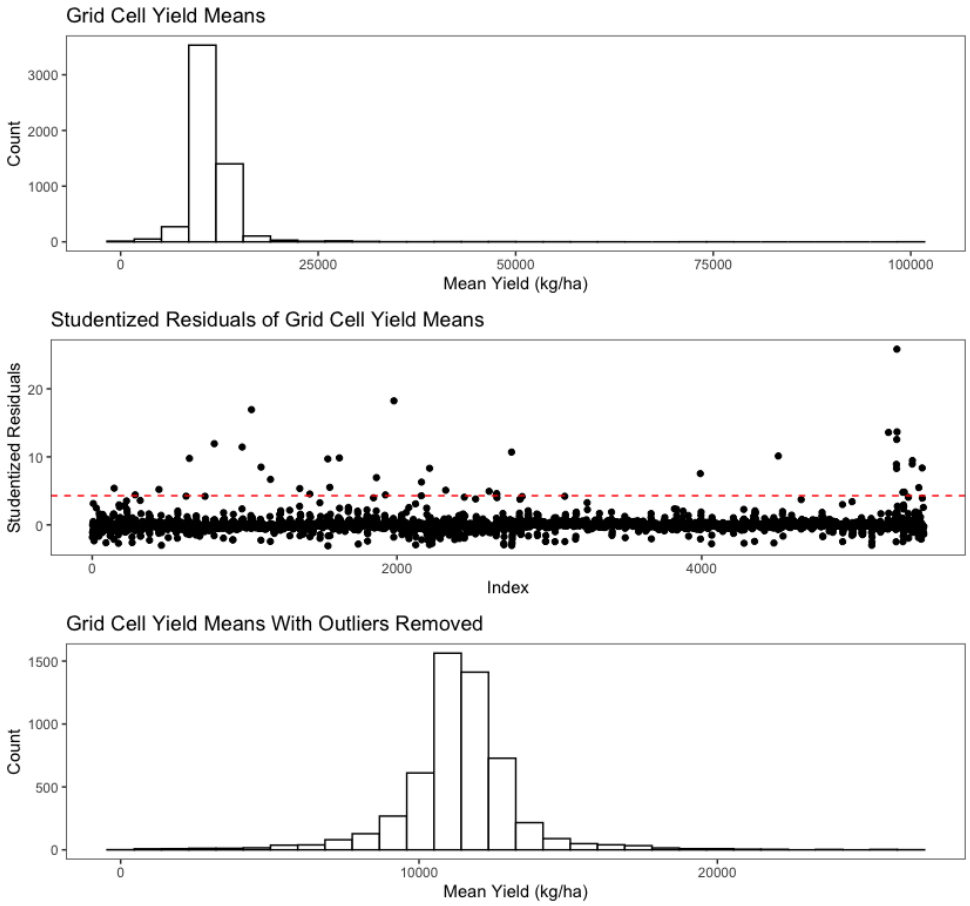
D Removing grid cells that contain data for more than one hybrid/variety



Due to the shape and features of each field, the planters occasionally deviated from a straight line path, resulting in data corresponding to more than on hybrid/variety falling within the same experimental unit grid cell. Using the unique IDs matching data points to grid cells, the hybrid/variety sown for each data point within each grid cell was identified in R. Cells that contained data points corresponding to more than one hybrid/variety were removed from the analysis. In the illustrated example, the cells outlined in black boxes would have been identified and removed from the analysis.

Figure S3.1 (Continued)

E Removing outlying grid cells based on the mean and variance of the as-applied planting rate and the yield



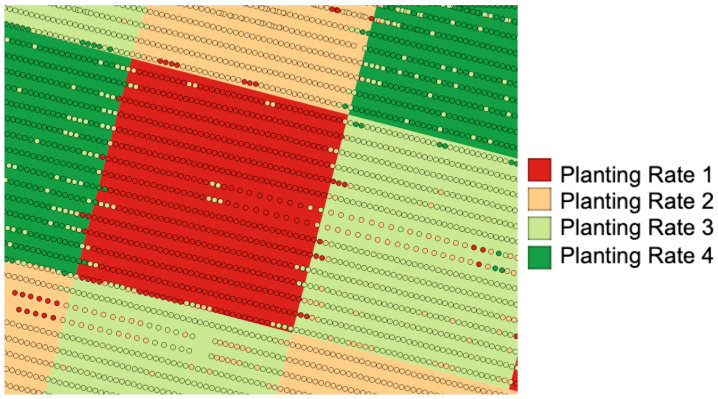
It was not uncommon for the planter and combine to record extreme values for the as-applied planting rate and the yield. To reduce the influence of these data points on the analysis, the mean and variance of the as-applied planting rate and yield for the data points within each grid cell were calculated in R. Significant outliers were identified using Studentized residuals (p-value < 0.05) with the lm() and rstudent() functions. Experimental grid cells were removed from the analysis if they were **Figure**

S3.1 (Continued)

identified as outliers with respect to the mean or the variance of the as-applied planting rate or the yield. The illustrated example shows the distribution of yield means before and after outlier removal.

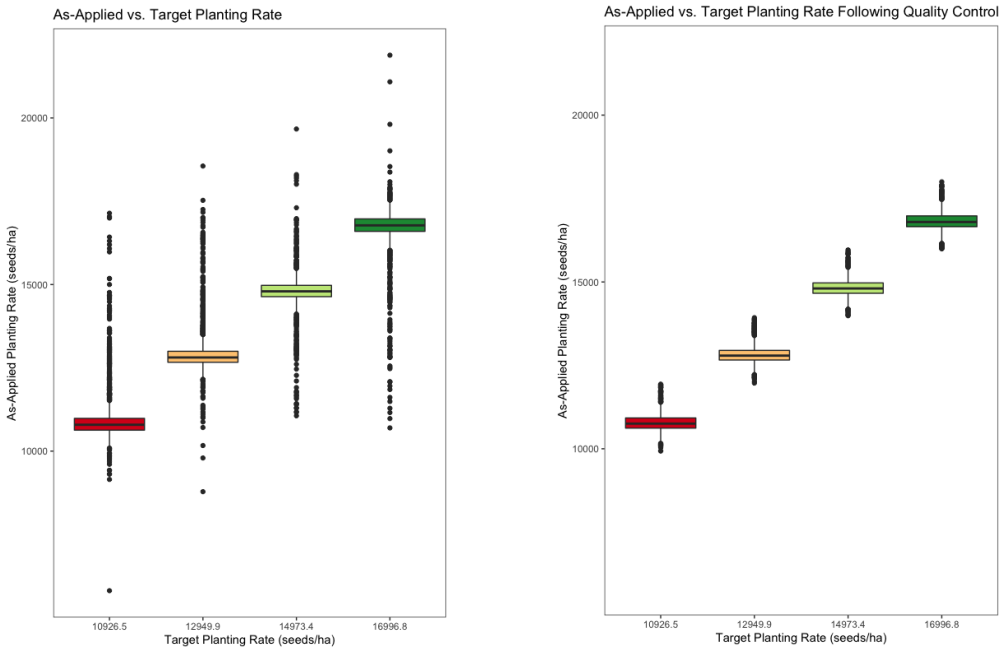
Figure S3.1 (Continued)

F Removing grid cells in which the as-applied planting rate did not correspond to the target rate



Occasionally, the as-applied planting rate did not match the target planting rate. In the illustrated example, the circles represent the as-applied data points and their colors indicate the as-applied rate. The blocks indicate the target planting rates.

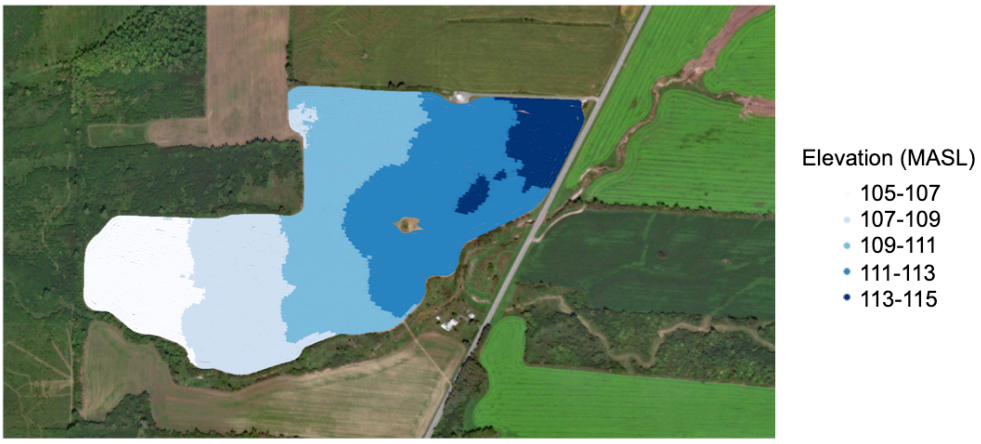
Figure S3.1 (Continued)



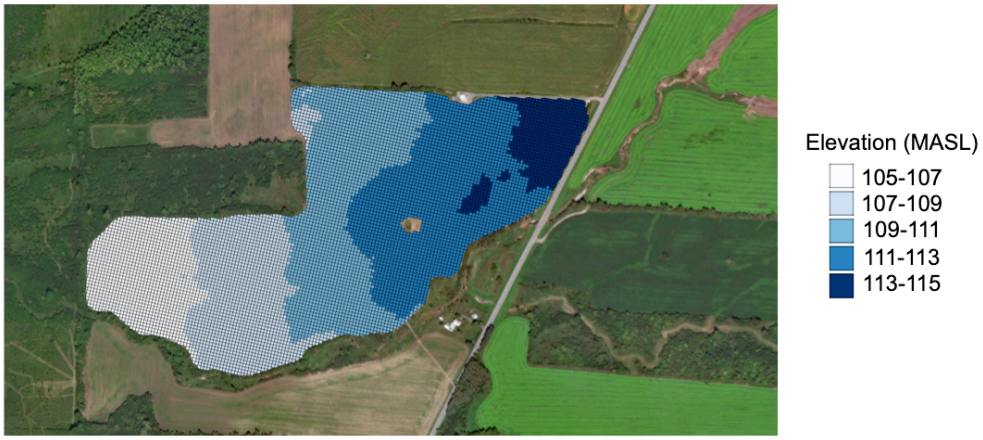
Experimental grid cells in which the as-applied planting rate data deviated substantially from the target planting rate were removed from the analysis. For corn, grid cells with as-applied planting rates that deviated from the target planting rate by > 1011.7 seeds ha^{-1} were removed from the analysis. For soybean, grid cells with as-applied planting rates that deviated from the target planting rate by > 8093.7 seeds ha^{-1} were removed from the analysis.

Figure S3.1 (Continued)

G Deriving the topographical variables from the as-applied planting rate layer

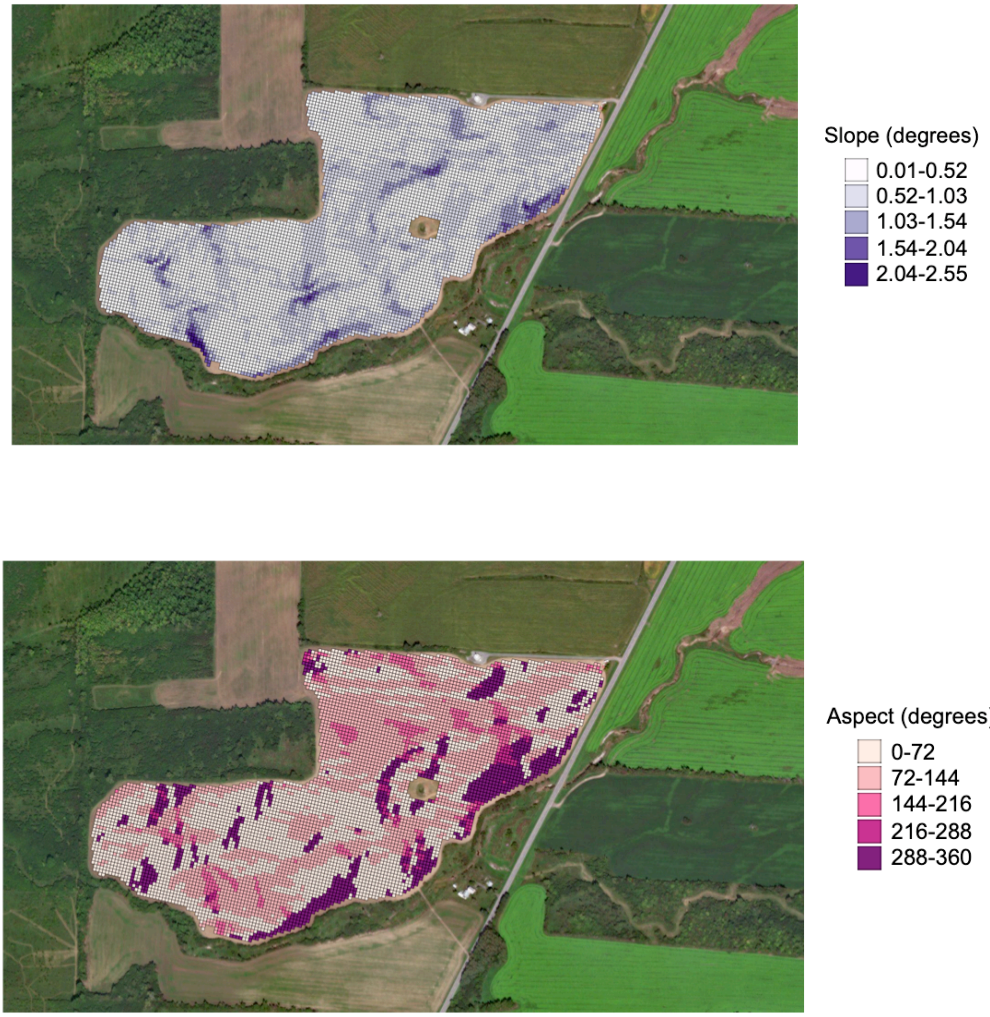


Elevation measurements were recorded with a GPS system aboard the planter.



Means of the elevation data points within each experimental unit grid cell were calculated using the “Geoprocessor” function of the “HTP Geoprocessor” plugin in QGIS such that a single value for elevation is assigned to each grid cell.

Figure S3.1 (Continued)



The slope and aspect for each experimental unit grid cell was calculated in R according to Burrough and McDonell (1998).

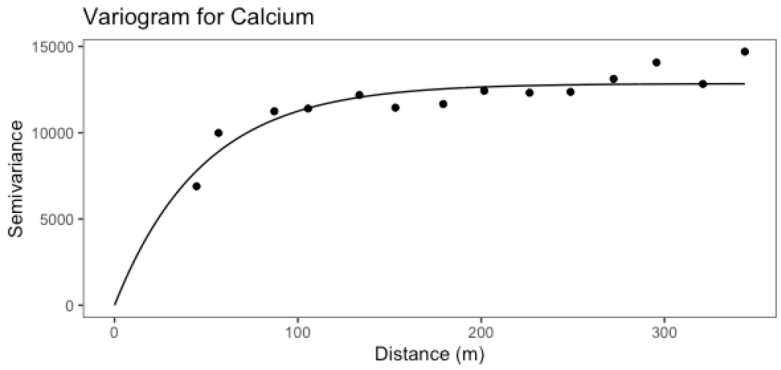
Figure S3.1 (Continued)

Curvature was calculated in R for each experimental unit grid cell according to Zeverbergen and Thorne (1987).

Figure S3.1 (Continued)

H Kriging the grid soil samples to the experimental unit grid

Soil grid samples were collected in a 0.20-ha grid pattern.



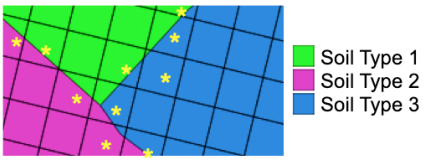
Using the “sp” and “gstat” packages in R, the variograms were fit using the fit.variogram() function for each of the 12 soil characteristics.

Figure S3.1 (Continued)

Centroids for each cell of the experimental unit grid were created using the “Polygon Centroids” function in QGIS. The krige() function from the “gstat” R packages was utilized to block krige the soil characteristics to grid. The “block” parameter set to the length and width of the grid cells.

Figure S3.1 (Continued)

I Setting the soil type to “missing” if more than one soil type was present within a grid cell



For experimental grid cells in which there was more than one soil type present, the soil type was set to missing. In the illustrated example, the yellow asterisks denote grid cells in which more than one soil type is present.

CHAPTER 4

CONCLUSION

The advent of digital technology has supplied plant breeders and growers with an array of novel technologies for understanding plant genome structure and function, measuring crop phenotypes in flux as the react to environmental stimuli, developing dense yield maps in growers' fields, and building high-resolution topographical and soil profiles of the agri-landscape. The rapid and exponential expansion of agricultural data with respect their size, scope, resolution, and diversity creates a myriad of new opportunities for plant breeders and growers to understand and optimize farming systems. However, the benefits of these new agricultural data resources – like data resources in medicine, business, and many other occupations – have not yet been fully realized. Several statistical and logistical challenges related to their theory and implementation are still to be addressed.

For example, with respect to breeding programs, aerial HTP platforms have the potential to dramatically increase the spatial and time scales of field-based phenotypic datasets available to plant breeders. This work and several other studies have demonstrated the potential for these traits to improve response to selection indirectly. However, a challenge of working with aerial HTP datasets is that they are often highly unbalanced. Weather, irrigation schedules, and equipment availability and function, among others, can hinder the acquisition of datasets in which all environments have been phenotyped at the same stage and the same number of times. This work presented a simple and practical approach for addressing the issue of unbalanced time-points, though it was not designed to effectively leverage all available information or to

accurately model finer-scale biological processes. There has been a growing interest among the plant breeding and crop modeling communities in merging crop growth modeling approaches with genetics to provide more accurate estimates of yield and other end-use traits; however, to date, most phenotypic datasets collected by plant breeders have been too sparse to effectively parameterize crop growth models. Increasing automation of aerial phenotyping technologies will enable data acquisition on a more frequent timescale within breeding programs. Integrating this information with existing biological knowledge encapsulated within crop growth models should be a future priority, as this approach may simultaneously improve our biological understand of crop growth and development and increase our selection accuracy within breeding programs.

Although selection on traits acquired with aerial or other types of HTP technologies may lead to a desired response in a correlated trait of interest, this work also demonstrated that such an approach might also drive unintended changes in the distribution of other correlated traits. A strong influence of days to heading on the relationship between aerial phenotypes and grain yield was observed in this work and has been shown in a number of other studies. As further developments in HTP technologies increase the availability of data at the early stages of breeding programs when seed stocks and/or resources are insufficient to reliably evaluate yield, it will be important for breeders to anticipate and address unintended directional responses to selection in phenotypes that may be confounded with the HTP traits, at least until those confounded traits can likewise be rapidly phenotyped and accounted for in multi-trait selection schemes. As shown, it will be important for these schemes to be able to account for the genetic correlations between traits.

Aerial HTP as a breeding technology is perhaps as far along the path from development to application as GS was five to ten years ago. Initial research efforts in

GS leveraged cross-validation to fine-tune the theoretical approaches and statistical modeling required to develop high quality genomic estimated breeding values. Once the greater community reached a certain level of consensus with respect to best practices, breeders and researchers began to pursue empirical studies to validate the established theory and address the logistical challenges of implementation. Over the past five to ten years, aerial HTP platforms have become smaller, cheaper, more reliable, and more functional. This and other recent works have contributed to improvements in post-processing and data extraction protocols for aerial imagery. The time is now to begin to empirically test aerial HTP and other ground-based systems within breeding programs for their ability to improve selection accuracy. This work provided a first step in that direction; however, no selections were made based fully or partially on the aerial HTP data. Thought and care should be allocated to balancing effective experimental designs with breeding program logistics when setting up empirical studies.

With respect to precision crop management, the New York Corn and Soybean Growers Association has made great strides in the evaluation of variable rate planting technologies by conducting, to the best of our knowledge, the largest on-farm study to-date of variable rate planting. This study greatly expanded the inference space for assessing the technology's logistical and economic potential. However, many challenges still remain. It is widely documented that different crop hybrids or varieties may exhibit differential responses to planting density. Therefore, to effectively optimize variable rate planting, new hybrids or varieties would need to undergo some years of testing to model their response to a range of planting densities. However, the current commercialized agricultural system in the United States operates such that new hybrids and varieties often have a short lifespan on the market before they are replaced by improved products. By the time sufficient testing has been performed to maximize

the economic benefits of variable rate planting for a particular hybrid or variety, it is possible that a new hybrid or variety may already have emerged on the market that offers a similar or superior marginal increase in yield when sown with a fixed rate as compared to the former hybrid or variety grown optimally under variable rate. Therefore, it may be important for researchers working to optimize variable rate planting to build partnerships with the seed industry leaders who are producing the products their stakeholders are purchasing. For example, variable rate testing could begin at the foundation seed stage prior to seed certification and release in order to assure that variable rate best practices can be established.

Technologies like variable rate planting are often plagued by genotype-by-environment interaction across sites and years. Several studies have suggested that variable rate planting may have limited potential for profitability given the wide spatial variation in the agricultural landscape across growers' fields and the weather variability from year-to-year. This worked showed that, as is often observed in plant breeding, prediction of yields based on topographical, soil, and planting rate variables across sites showed low levels of accuracy. However, accuracies were higher when predicting training and testing across years at the same location. It is likely that variable rate planting designs developed with data recorded in the field which is to be predicted will be more accurate than building designs based on models trained with others locations' data. It may therefore be most advantageous for the community to adopt a model of "growers as researchers." At the most basic level, growers could work with public sector scientists to provide on-farm data for their traditional practices in order to aid the development of precision agriculture technologies. With this data, researchers could develop site-specific strategies with a greater outlook for profitability. At a higher level, growers could implement experimental designs on their farms as demonstrated by the New York Corn and Soybean Growers Association.

Facilitated by grower organizations and public sector researchers, these experiments could produce data to develop more advanced technologies that could likewise produce detailed, site-specific strategies for planting rate and other management practices. The New York Corn and Soybean Growers Association has set a commendable example of what this type of model may look like in its early stages. The project could further benefit from additional participation, investment, and collaboration.

The works presented herein provide some of the foundational theory and logistical aspects of scaling novel approaches to GS, HTP and spatial analysis, though these are in many ways the first steps towards their successful application in plant breeding and crop management. There are many more opportunities to further explore theoretical research, empirical studies, and on-farm collaborations in order to fully maximize the benefits of these emerging technologies and data resources.