

# Multitrait, Random Regression, or Simple Repeatability Model in High-Throughput Phenotyping Data Improve Genomic Prediction for Wheat Grain Yield

Jin Sun, Jessica E. Rutkoski, Jesse A. Poland, José Crossa, Jean-Luc Jannink, and Mark E. Sorrells\*

## Abstract

High-throughput phenotyping (HTP) platforms can be used to measure traits that are genetically correlated with wheat (*Triticum aestivum* L.) grain yield across time. Incorporating such secondary traits in the multivariate pedigree and genomic prediction models would be desirable to improve indirect selection for grain yield. In this study, we evaluated three statistical models, simple repeatability (SR), multitrait (MT), and random regression (RR), for the longitudinal data of secondary traits and compared the impact of the proposed models for secondary traits on their predictive abilities for grain yield. Grain yield and secondary traits, canopy temperature (CT) and normalized difference vegetation index (NDVI), were collected in five diverse environments for 557 wheat lines with available pedigree and genomic information. A two-stage analysis was applied for pedigree and genomic selection (GS). First, secondary traits were fitted by SR, MT, or RR models, separately, within each environment. Then, best linear unbiased predictions (BLUPs) of secondary traits from the above models were used in the multivariate prediction models to compare predictive abilities for grain yield. Predictive ability was substantially improved by 70%, on average, from multivariate pedigree and genomic models when including secondary traits in both training and test populations. Additionally, (i) predictive abilities slightly varied for MT, RR, or SR models in this data set, (ii) results indicated that including BLUPs of secondary traits from the MT model was the best in severe drought, and (iii) the RR model was slightly better than SR and MT models under drought environment.

## Core Ideas

- HTP platforms used to measure secondary traits across time
- Longitudinal data of secondary traits evaluated by SR, MT, and RR models, separately
- BLUPs of secondary traits used in the multivariate pedigree and genomic prediction
- Grain yield predictive ability was improved by 70%

**W**ITH the advent of next-generation sequencing technology, GS becomes feasible, making it possible to accelerate breeding cycles over traditional approaches. Genomic selection aims to take advantage of low-cost, genome-wide molecular markers to increase genetic gain of complex quantitative traits (Poland et al., 2012), and it has been applied in wheat (Rutkoski et al., 2011, 2012; Poland et al., 2012). On the other hand, there is increasing focus on HTP platforms (Araus and Cairns, 2014) that are

J. Sun, J.E. Rutkoski, J.L. Jannink, M.E. Sorrells, Dep. of Plant Breeding and Genetics, Cornell Univ., Ithaca, NY 14853, USA; J.E. Rutkoski, J. Crossa, CIMMYT, Km. 45, Carretera México-Veracruz, El Batán, Texcoco CP 56237, Mexico; J.A. Poland, Dep. of Plant Pathology and Dep. of Agronomy, Kansas State Univ., Manhattan, KS 66506, USA; J.L. Jannink, USDA-ARS R.W. Holley Center for Agriculture and Health, Ithaca, NY 14853, USA. Received 16 Nov. 2016. Accepted 6 Feb. 2017. \*Corresponding author (mes12@cornell.edu). Assigned to Associate Editor Nicholas Tinker.

**Abbreviations:** BLUP, best linear unbiased prediction; BV, bivariate prediction; CT, canopy temperature; GNDVI, green NDVI; GS, genomic selection; HTP, high-throughput phenotyping; ISE, indirect selection efficiency; MT, multitrait; MV1, first multivariate prediction; MV2, second multivariate prediction; NDVI, normalized difference vegetation index; RNDVI, red NDVI; RR, random regression; SNP, single-nucleotide polymorphism; SR, simple repeatability; UV, univariate prediction.

Published in Plant Genome  
Volume 10. doi: 10.3835/plantgenome2016.11.0111

© Crop Science Society of America  
5585 Guilford Rd., Madison, WI 53711 USA  
This is an open access article distributed under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

capable of measuring large-scale, high-density phenotyping across time at high accuracy and low labor intensity (Yang et al., 2014). If the traits collected from HTP platforms are genetically correlated with the primary trait, such traits could be considered as secondary traits to improve rates of genetic gain for the primary trait in GS. In addition, using secondary traits collected from HTP platforms would also be useful to predict primary trait at early growth stages, as they could be phenotyped ahead of the primary trait like grain yield.

Grain yield is a complex quantitative trait that is controlled by many genes and is influenced by environment (Narjesi et al., 2015). Traits like CT and NDVI can be measured from HTP platforms continuously across plant growth stages. They are genetically correlated with grain yield and have been applied to predict grain yield in previous studies (Labus et al., 2002; Mason and Singh, 2014; Quarmbly et al., 1993; Rees et al., 1993). The negative correlation between CT and grain yield can be used as a selection criterion in wheat breeding (Rees et al., 1993; Mason and Singh, 2014), particularly under heat-stressed environments (Mason and Singh, 2014). The green NDVI (GNDVI) and red NDVI (RNDVI) are vegetation indices calculated by the difference between near-infrared reflectance and green–red reflectance (Gitelson et al., 1996; Tucker, 1979), and wheat grain yield could be well estimated before harvest using NDVI based on a linear relationship between yield and the index (Labus et al., 2002; Quarmbly et al., 1993). Multivariate pedigree and genomic prediction models have been shown to improve prediction accuracy when there is a strong correlation between secondary traits and a primary trait (Pszczola et al., 2013; Guo et al., 2014). Rutkoski et al. (2016) found that including CT and NDVI as secondary traits for grain yield in a multivariate prediction model could improve prediction accuracies vs. a univariate model.

In GS, it is more computationally efficient and simpler to use two-stage analysis than a one-stage analysis. Therefore, prior to using the secondary traits in prediction models, secondary traits could be first fit with the models that are able to handle data from HTP platforms. Canopy temperature and NDVI from HTP platforms are measured at many time points throughout the growth cycle and can be considered as longitudinal data. Such data sets are common in animal breeding for traits such as body weight of cattle growth. Models generally applied to longitudinal data in animal breeding include the SR (also named repeatability model in animal breeding), MT, and RR models. Rutkoski et al. (2016) used the SR model for secondary traits collected across time points by considering each time point within a growth stage (either vegetative or grain filling) as a repetitive collection for the same trait and incorporated best linear unbiased estimates of secondary traits from the SR model into multivariate pedigree and genomic prediction models. However, the SR model assumes constant variance and correlation at or between measurement dates, which may not be realistic for longitudinal data collected

at different time points especially when time points span growth stages (Meyer and Hill, 1997). In the MT model, secondary traits at different time points are considered as multiple response variables for each line. However, a high correlation between consecutive measurements and computational requirements can restrict the application of the MT model (Speidel, 2011). The RR model has been applied to capture the change of a trait continually over growth stages with fewer parameters than the MT model in animal breeding. Furthermore, the assumptions for constant variances and correlations are not required in the RR model compared with the SR model (Meyer, 2000). A covariance at or between each time point can also be fitted in the RR model using different functions, typically Legendre polynomials (Meyer, 2005). Nonetheless, it has been observed that high-degree Legendre polynomials rapidly changed at the extremes, causing poor estimates of genetic parameters and variance components (Boligon et al., 2012; Meyer, 2005; Misztal, 2006). Compared with Legendre polynomials, spline functions with fractional polynomials are less prone to those problems because the growth curve is joined smoothly by short segments and each segment is only determined by each coefficient of the spline function (Boligon et al., 2012; Meyer, 2005; Misztal, 2006).

The objectives of this study were to (i) evaluate statistical models that could efficiently handle the longitudinal data for secondary traits from HTP platforms by comparing SR, MT and RR models and (ii) test the impact of the proposed models for the secondary traits based on their heritabilities, genetic correlations with grain yield, and genomic predictive abilities for grain yield.

## Materials and Methods

### Phenotypic Data

In our study, the data set was the same as reported in Rutkoski et al. (2016). We used 630 wheat inbred breeding lines for subsequent analysis, and they were from the 2013–2014 growing season at the International Wheat and Maize Improvement Center (CIMMYT) in Obregon, Mexico. Those lines were grouped into 21 trials, and each trial includes 28 lines and two checks in an  $\alpha$ -lattice design with three replicates and six blocks. The trials were planted in five diverse environments: optimal, early heat, late heat, drought, and severe drought. The sowing and irrigation conditions in each environment have been described in Rutkoski et al. (2016). High-throughput phenotyping data (secondary traits: CT and NDVI) were collected by the hyperspectral and thermal camera in an aircraft over wheat growth stages, and grain yield for those lines were evaluated. The number of dates available for each secondary trait varied in each environment. There were eight, seven, three, five, and six time points available for optimal, early heat, late heat, drought, and severe drought, respectively, and all time points were considered in either vegetative or grain filling stages.

## Genotyping

Genome-wide genotyping of lines were based on genotyping-by-sequencing. Single-nucleotide polymorphism (SNP) markers were filtered, resulting in 12,083 SNPs available for 557 individuals. The criteria for filtering were as follows: the markers were removed if >80% of the individuals had missing data for a SNP, or if >20% of individuals were heterozygous for a SNP, and lines had >80% missing markers were removed. In addition, markers were also filtered for minor allele frequency <0.01, and missing data were imputed based on the mean of markers (Rutkoski et al., 2016).

## Genetic Value Prediction

Two-step GS was applied in this study. In the first step, BLUPs of 630 lines were predicted for grain yield and secondary traits, separately. The BLUPs of secondary traits were predicted from the first two replicates, and the BLUPs of grain yield were predicted using the third replicate. Using BLUPs of secondary traits and grain yield in different replicates for model training and validation ensures that the genomic prediction accuracy for grain yield from secondary traits will not be inflated by the sharing of specific–environment effects.

For grain yield, BLUPs of each genotype were predicted based on a simple mixed model using the data from the third replicate within each environment, the matrix notation is as follows:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{g} + \mathbf{W}\mathbf{t} + \mathbf{Q}\mathbf{p} + \mathbf{e} \quad [1]$$

where  $\mathbf{y}$  is the vector of observations for grain yield;  $\mathbf{X}$ ,  $\mathbf{Z}$ ,  $\mathbf{W}$ , and  $\mathbf{Q}$  are incidence matrices corresponding to the fixed effect ( $\mathbf{b}$ ), random genetic effect ( $\mathbf{g}$ ), random environmental trial effect ( $\mathbf{t}$ ), and random environmental block effects ( $\mathbf{p}$ ); and  $\mathbf{e}$  is the random residual errors. The variance and covariance structures are based on the following assumptions:  $\mathbf{g} \sim N(0, \mathbf{I}\sigma_g^2)$ ,  $\mathbf{t} \sim N(0, \mathbf{I}\sigma_t^2)$ ,  $\mathbf{p} \sim N(0, \mathbf{I}\sigma_p^2)$ , and  $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$ ;  $\sigma_g^2$  is the genetic variance;  $\sigma_t^2$  and  $\sigma_p^2$  are environmental variances;  $\sigma_e^2$  is the residual variance; and  $\mathbf{I}$  is the identity matrix.

For secondary traits, CT and NDVI, BLUPs of each genotype were predicted by fitting three models (SR, MT, or RR spline models) separately using the first two replicates. Since available time points varied for each secondary trait under each environment, three models were also fit for each trait within each environment.

## Simple Repeatability Model

Data collected at each time point in the SR model were considered as repeated records of the same trait for each line. The SR model was the same as Eq. [1] with the fixed effect for time point and additional random effect for replicate, the matrix format is as follows:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{g} + \mathbf{W}\mathbf{t} + \mathbf{S}\mathbf{r} + \mathbf{Q}\mathbf{p} + \mathbf{e} \quad [2]$$

where  $\mathbf{y}$  is the vector of observations for secondary traits;  $\mathbf{X}$ ,  $\mathbf{Z}$ ,  $\mathbf{W}$ ,  $\mathbf{S}$ , and  $\mathbf{Q}$  are incidence matrices

corresponding to the fixed effect for time point ( $\mathbf{b}$ ), random genetic effect ( $\mathbf{g}$ ), random environmental effect for trial ( $\mathbf{t}$ ), random environmental effect for replicate ( $\mathbf{r}$ ), and random environmental effect for block ( $\mathbf{p}$ ); and  $\mathbf{e}$  is the random residual errors. The variance and covariance structures are based on the following assumptions:  $\mathbf{g} \sim N(0, \mathbf{I}\sigma_g^2)$ ,  $\mathbf{t} \sim N(0, \mathbf{I}\sigma_t^2)$ ,  $\mathbf{r} \sim N(0, \mathbf{I}\sigma_r^2)$ ,  $\mathbf{p} \sim N(0, \mathbf{I}\sigma_p^2)$ , and  $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$ ;  $\sigma_g^2$  is the genetic variance;  $\sigma_t^2$ ,  $\sigma_r^2$ , and  $\sigma_p^2$  are environmental variances;  $\sigma_e^2$  is the residual variance; and  $\mathbf{I}$  is the identity matrix.

## Multitrait Model

For the MT model, data collected at each time point was considered as a single trait. The MT model in matrix form was expressed as the following:

$$\begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_n \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{X}_n \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_n \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{Z}_n \end{bmatrix} \begin{bmatrix} \mathbf{g}_1 \\ \vdots \\ \mathbf{g}_n \end{bmatrix} + \begin{bmatrix} \mathbf{W}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{W}_n \end{bmatrix} \begin{bmatrix} \mathbf{t}_1 \\ \vdots \\ \mathbf{t}_n \end{bmatrix} + \begin{bmatrix} \mathbf{S}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{S}_n \end{bmatrix} \begin{bmatrix} \mathbf{r}_1 \\ \vdots \\ \mathbf{r}_n \end{bmatrix} + \begin{bmatrix} \mathbf{Q}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{Q}_n \end{bmatrix} \begin{bmatrix} \mathbf{p}_1 \\ \vdots \\ \mathbf{p}_n \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_n \end{bmatrix} \quad [3]$$

where  $\mathbf{y}_i$  is the vector of observations for the secondary trait at  $i$ th time point;  $\mathbf{X}_i$ ,  $\mathbf{Z}_i$ ,  $\mathbf{W}_i$ ,  $\mathbf{S}_i$ , and  $\mathbf{Q}_i$  are incidence matrices relating the observations in  $\mathbf{y}_i$  to the fixed effects in  $\mathbf{b}_i$ , random genetic effects in  $\mathbf{g}_i$ , random environmental effects (trial, replicate, and block) in  $\mathbf{t}_i$ ,  $\mathbf{r}_i$ , and  $\mathbf{p}_i$ , respectively, at  $i$ th time point; and  $n$  is number of time points for each secondary trait in each environment. The variance components are defined as

$$\begin{bmatrix} \mathbf{g}_1 \\ \vdots \\ \mathbf{g}_n \end{bmatrix} \sim N(0, \mathbf{I} \otimes \mathbf{K}_g), \quad \begin{bmatrix} \mathbf{t}_1 \\ \vdots \\ \mathbf{t}_n \end{bmatrix} \sim N(0, \mathbf{I} \otimes \mathbf{K}_t), \quad \begin{bmatrix} \mathbf{r}_1 \\ \vdots \\ \mathbf{r}_n \end{bmatrix} \sim N(0, \mathbf{I} \otimes \mathbf{K}_r),$$

$$\begin{bmatrix} \mathbf{p}_1 \\ \vdots \\ \mathbf{p}_n \end{bmatrix} \sim N(0, \mathbf{I} \otimes \mathbf{K}_p), \quad \begin{bmatrix} \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_n \end{bmatrix} \sim N(0, \mathbf{I} \otimes \mathbf{R}), \quad \text{where } \mathbf{K}_g \text{ is a}$$

covariance matrix of random genetic effects,  $\mathbf{K}_t$ ,  $\mathbf{K}_p$ , and  $\mathbf{K}_p$  are covariance matrices of random environmental effects;  $\mathbf{R}$  is a covariance matrix of random residual effects;  $\mathbf{I}$  is an identity matrix with different orders for different random effects; and  $\otimes$  denotes the Kronecker product. In addition,  $\mathbf{K}_g$  and  $\mathbf{R}$  are assumed unstructured variance–covariance structures in the model, and  $\mathbf{K}_t$ ,  $\mathbf{K}_p$ , and  $\mathbf{K}_p$  are diagonal variance–covariance structures. The overall BLUP for each genotype in each environment from MT model were averaged over time points in each environment.

## Random Regression Model using Cubic Smoothing Spline

Let  $y_{ijklm}$  be the observations for secondary trait for  $j$ th line,  $k$ th trial,  $l$ th replicate, and  $m$ th block at time point  $i$ . The following model is derived from White et al. (1999), which is equivalent to a RR model (Meyer, 1998, 2005), and additional permanent random environmental effects was included. The model was fitted using a cubic smoothing spline in ASReml-R (Butler et al., 2009). A spline is a curve that is joined continuously by multiple polynomial segments, and each joint is referred to as a knot (Meyer, 2005). A cubic smoothing spline consists of piecewise cubic functions, and a roughness penalty is imposed to determine the balance between smoothness and reliability of the data (White et al., 1999). Green and Silverman (1994) discussed details about cubic smoothing splines:

$$y_{ijklm} = b_0 + b_1 t_{ijklm} + b_{j0} + b_{j1} t_{ijklm} + \sum_{n=2}^{q-1} v_{jn} z_n(t_{ijklm}) + \sum_{n=2}^{q-1} v_{jn} z_n(t_{ijklm}) + b_{k0} + b_{k1} t_{ijklm} + \sum_{n=2}^{q-1} v_{kn} z_n(t_{ijklm}) + b_{l0} + b_{l1} t_{ijklm} + \sum_{n=2}^{q-1} v_{ln} z_n(t_{ijklm}) + b_{m0} + b_{m1} t_{ijklm} + \sum_{n=2}^{q-1} v_{mn} z_n(t_{ijklm}) + \varepsilon_{ijklm} \quad [4]$$

where  $t_{ijklm}$  is the time point  $i$  for line  $j$  at trial  $k$ , replicate  $l$ , and block  $m$  and  $b_0$  (intercept) and  $b_1$  (slope) are the general linear regression over time  $t_{ijklm}$  as fixed. The rest of terms in Eq. [4] are random:  $b_{j0}$  (line) and  $b_{j1}$  (line  $\times$  linear) capture the deviation from general linear regression for line  $j$ , similar for  $b_{k0}$  (trial);  $b_{k1}$  (trial  $\times$  linear) for trial  $k$ ;  $b_{l0}$  (replicate) and  $b_{l1}$  (replicate  $\times$  linear) for replicate  $l$ ;  $b_{m0}$  (block) and  $b_{m1}$  (block  $\times$  linear) for block  $m$ ;

$\sum_{n=2}^{q-1} v_{jn} z_n(t_{ijklm})$  is a mean spline deviation,  $z_n(t_{ijklm})$  is the function of time points as covariables, cubic smooth spline function was implemented in this model,  $v_n$  is the regression coefficient for each spline function  $z_n(t_{ijklm})$ , and  $q$  is the number of regression coefficients (number of knots for spline in this model) fitted for genetic and

environmental effects;  $\sum_{n=2}^{q-1} v_{jn} z_n(t_{ijklm})$  (line  $\times$  spline)

and other similar terms,  $\sum_{n=2}^{q-1} v_{kn} z_n(t_{ijklm})$  (trial  $\times$  spline),

$\sum_{n=2}^{q-1} v_{ln} z_n(t_{ijklm})$  (replicate  $\times$  spline) and  $\sum_{n=2}^{q-1} v_{mn} z_n(t_{ijklm})$

(block  $\times$  spline), are deviation from the mean spline for additive genetic effect of line  $j$  and permanent environmental effects including trial  $k$ , replicate  $l$ , and block  $m$  for line  $j$ ; and  $\varepsilon_{ijklm}$  is the residual error term. In this model, number of knots was the same as the number of time points for each secondary trait under each environment. Detailed information about this model can be found in White et al. (1999).

The matrix notation for RR model is (Mrode, 2005):

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zg} + \mathbf{Wt} + \mathbf{Sr} + \mathbf{Qp} + \mathbf{e} \quad [5]$$

where  $\mathbf{y}$  is the vector of observations for secondary traits,  $\mathbf{X}$  is the incidence matrix corresponding to fixed effects and fixed regression coefficients,  $\mathbf{b}$  is the vector for fixed effect and fixed regression. The matrices  $\mathbf{Z}$ ,  $\mathbf{W}$ ,  $\mathbf{S}$ , and  $\mathbf{Q}$  are covariable matrices relating to random genetic, random environmental trial, replicate, and block effects; and  $\mathbf{g}$ ,  $\mathbf{t}$ ,  $\mathbf{r}$ , and  $\mathbf{p}$  are vectors of random regressions for genetic and environmental effects. The variance components are assumed as:  $\mathbf{g} \sim N(0, \mathbf{I} \otimes \mathbf{K}_g)$ ,  $\mathbf{t} \sim N(0, \mathbf{I} \otimes \mathbf{K}_t)$ ,  $\mathbf{r} \sim N(0, \mathbf{I} \otimes \mathbf{K}_r)$ ,  $\mathbf{p} \sim N(0, \mathbf{I} \otimes \mathbf{K}_p)$ ,  $\mathbf{e} \sim N(0, \mathbf{I} \otimes \mathbf{K}_e)$ , where  $\mathbf{K}_g$ ,  $\mathbf{K}_t$ ,  $\mathbf{K}_r$ , and  $\mathbf{K}_p$  are covariance matrices (of the order equal to the order of the spline polynomial fitted) between random regression coefficients for random genetic and random environmental (trial, replicate, and block) effects, respectively;  $\mathbf{I}$  is the identity with different orders corresponding to genetic and environmental effects, and  $\otimes$  denotes the Kronocker product.

The predict function implemented in ASReml-R was used to calculate the BLUP for each line at each time point. The overall BLUP for each line was averaged over time points in each environment, which was the same approach used for the MT model.

## Heritability, Correlation, and Indirect Selection Efficiency

Variance components for narrow-sense heritability for each secondary trait and grain yield in each environment were estimated using the following model:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zg} + \mathbf{e} \quad [6]$$

where  $\mathbf{y}$  is the BLUPs of genotypes from SR, MT, or RR spline models for secondary traits or BLUPs of genotypes from the mixed model Eq. [1] for grain yield;  $\mathbf{X}$  and  $\mathbf{Z}$  are incidence matrices corresponding to the fixed effect ( $\mathbf{b}$ ) and random genetic effect ( $\mathbf{g}$ ); and  $\mathbf{e}$  is the random residual errors. The variance and covariance structures are based on the following assumptions:  $\mathbf{g} \sim N(0, \mathbf{G}\sigma_a^2)$ , where  $\mathbf{G}$  is the genomic relationship matrix, or  $\mathbf{g} \sim N(0, \mathbf{A}\sigma_a^2)$ , where  $\mathbf{A}$  is the pedigree relationship matrix and  $\sigma_a^2$  is the additive genetic variance;  $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$ , where  $\sigma_e^2$  is the residual variance; and  $\mathbf{I}$  is the identity matrix. Narrow-

sense heritability was calculated as  $h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2}$ .

Variance and covariance components for correlations were estimated using the multivariate model (Eq. [5]) in Rutkoski et al. (2016) in each environment. The response variables ( $y$ ) were the BLUPs of genotypes from SR, MT, or RR models for secondary traits and BLUPs of genotypes from the mixed model (Eq. [1]) for grain yield. Genetic correlations between secondary traits and grain yield were calculated as follows:

$$r_{g(\text{ST}, \text{GRYLD})} = \frac{\text{cov}_g(\text{ST}, \text{GRYLD})}{\sqrt{\text{var}_g(\text{ST}) \text{var}_g(\text{GRYLD})}}$$

where  $r_{g(\text{ST},\text{GRYLD})}$  is the genetic correlation between a secondary trait (either CT or NDVI) and grain yield,  $\text{var}_g(\text{ST})$  and  $\text{var}_g(\text{GRYLD})$  are the genetic variances of secondary trait and grain yield, respectively, and  $\text{cov}_g(\text{ST},\text{GRYLD})$  is the genetic covariance between a secondary trait and grain yield. Phenotypic correlations between secondary traits and grain yield were also calculated based on Pearson correlations.

Indirect selection efficiency (ISE) was estimated based on Falconer and Mackay (1996) under the assumption of the same intensity for both direct and indirect selection:

$$\text{ISE} = \sqrt{h_{\text{ST}}^2 / h_{\text{GRYLD}}^2} \times \left| r_{g(\text{ST},\text{GRYLD})} \right|$$

where  $h_{\text{ST}}^2$  is the narrow-sense heritability for a secondary trait (CT or NDVI),  $h_{\text{GRYLD}}^2$  is the narrow-sense heritability for grain yield, and  $r_{g(\text{ST},\text{GRYLD})}$  is the genetic correlation between a secondary trait and grain yield estimated either using genomic or pedigree based relationships.

### Genomic Prediction and Cross Validation

In the second step of GS, the BLUPs of 557 individuals from the SR, MT, or RR model for secondary traits or from the mixed model (Eq. [1]) for grain yield were used as dependent variables in the genomic prediction modeling. As the lines in this data set are replicated the same number of times in each environment, differential shrinkage of the BLUPs used as the dependent variable for genomic prediction is not an issue in this case. Four different prediction models (Supplemental Fig. S1) were used to compare SR, MT, and RR models and to check the efficiency of genomic and pedigree prediction with secondary traits.

The univariate prediction (UV) model with grain yield only is the same as the model expressed in Eq. [6], where  $\mathbf{y}$  is the BLUPs of genotypes from the mixed model (Eq. [1]) for grain yield only.

Two multivariate genomic prediction models were applied to identify the improvement of genomic prediction for grain yield with additional secondary traits in the model fitting. For the first multivariate prediction (MV1) model, three secondary traits were included in both training and testing population (Supplemental Fig. S1). For the second multivariate prediction (MV2) model, three secondary traits were only included in the training population (Supplemental Fig. S1). The model was as follows:

$$\begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_n \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{X}_n \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_n \end{bmatrix} + \begin{bmatrix} \mathbf{Z}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{Z}_n \end{bmatrix} \begin{bmatrix} \mathbf{g}_1 \\ \vdots \\ \mathbf{g}_n \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_n \end{bmatrix} \quad [7]$$

where  $n$  is the number of traits (here  $n = 4$ , including three secondary traits and grain yield);  $\mathbf{y}_i$  are BLUPs of genotypes from SR, MT, or RR spline models for secondary traits and BLUPs of genotypes from the mixed model Eq. [1] for grain yield;  $\mathbf{X}_i$  and  $\mathbf{Z}_i$  are the fixed and random effects design matrix, individually,  $\mathbf{b}_i$ ,  $\mathbf{g}_i$ , and  $\mathbf{e}_i$  are vectors of fixed effects, random genetic, and residual effects for

each trait, respectively. Variance components were estimated by assuming  $\begin{bmatrix} \mathbf{g}_1 \\ \vdots \\ \mathbf{g}_n \end{bmatrix} \sim N(0, \mathbf{G} \otimes \mathbf{H})$  or  $\begin{bmatrix} \mathbf{g}_1 \\ \vdots \\ \mathbf{g}_n \end{bmatrix} \sim N(0, \mathbf{A} \otimes \mathbf{H})$ ,

where  $\mathbf{G}$  or  $\mathbf{A}$  is the genomic or pedigree relationship matrices, and  $\mathbf{H}$  is the variance–covariance matrix for traits; and  $\begin{bmatrix} \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_n \end{bmatrix} \sim N(0, \mathbf{I} \otimes \mathbf{R})$ , where  $\mathbf{I}$  is an identity matrix,

and  $\mathbf{R}$  is the residual variance–covariance matrix between traits. Both  $\mathbf{H}$  and  $\mathbf{R}$  are assumed as unstructured.

In addition to those prediction models, the improvement of predictive ability from a single secondary trait was also tested. Bivariate prediction (BV) model, which included additional single secondary trait in both training and testing populations (Supplemental Fig. S1), was the same as MV model (Eq. [7]). The number of traits ( $n = 2$ ) includes grain yield and one of the secondary traits in the model.

Five-fold cross-validation was applied for all predictions. For each fold, predictive abilities were calculated as the Pearson correlation between BLUPs of grain yield from mixed model (Eq. [1]) and estimated breeding values of grain yield from the above prediction models in the testing population based on either genomic or pedigree relationship matrix. Cross-validation was conducted within each environment.

### Software Package

Data analysis was implemented in the R environment (R Development Core Team, 2010), and all models were fitted in ASReml-R (Butler et al., 2009). Genomic relationship matrix was calculated according to Eq. [15] in Endelman and Jannink (2012) using the R package rrBLUP (Endelman, 2011). The pedigree relationship matrix was estimated as  $2 \times$  the coefficient of parentage using the International Crop Information System.

## Results

### Heritability

The heritability estimates for grain yield (Table 1) were lower than those of NDVI (Table 2) in all environments, whereas they were higher than the heritability estimates of CT (Table 2) in late heat and severe drought environments. Among five environments, relatively high heritabilities for secondary traits and relatively low heritability for grain yield were both observed in the optimal environment. Among SR, MT, and RR models, higher heritability estimates for secondary traits were observed using BLUPs from the MT model than the SR and RR models under most environments except severe drought. In addition, heritability estimates using the genomic relationship matrix were higher than those using the pedigree relationship matrix for both grain yield and secondary traits under all five environments.

## Correlations Between Secondary Traits and Grain Yield

Correlations (Table 3) between secondary traits and grain yield varied across environments. Canopy temperature showed a higher correlation with grain yield than NDVI in most environments except severe drought, where NDVI had higher correlation with grain yield than CT. Among five environments, the highest correlations between secondary traits and grain yield were observed in the late heat for CT and in the severe drought for NDVI. Among SR, MT, and RR models, higher correlations between secondary traits and grain yield were observed using BLUPs from SR and RR models than MT model under most environments except severe drought where MT model was better than the other two. For CT, higher correlations with grain yield using the genomic rather than the pedigree relationship matrix were observed under late heat and drought, and for NDVI, this was observed in late heat and severe drought environments.

## Predictive Abilities

Four different prediction models (Supplemental Fig. S1) were tested in this data set. The multivariate MV1 model outperformed the MV2 model and the UV model; however, the MV2 model showed similar predictive ability as UV even though secondary traits were included in the training population for the MV2 model. From the comparison between prediction models UV and MV1 (Fig. 1; above MV1), the average improvement of predictive ability for grain yield over UV was 70% based on either genomic or pedigree relationship matrix. Besides, the predictive ability from the BV model with a single secondary trait (Fig. 2) was superior to the UV model as well, and including CT in BV provided better predictive ability for grain yield than NDVI in all environments except drought-stressed environments. Additionally, positive relationships were observed between predictive ability gained from a single secondary trait and ISE of single secondary trait from SR, MT, or RR models (Fig. 3).

For secondary traits included in either multivariate or BV models, we used BLUPs of secondary traits from SR, MT, or RR models, individually. Based on the predictive ability from the multivariate MV1 model with all three secondary traits, among SR, MT, and RR models, MT and SR models performed slightly better than RR model in the heat-stressed environments (Fig. 1; above MV1). However, predictive ability based on BLUPs from the MT and RR models were superior to SR model in drought-stressed environments, where MT model was the best in severe drought environment and RR model was slightly better than others in the drought environment (Fig. 1; above MV1).

## Discussion

### Secondary Traits

#### *Heritability and Correlation*

It was reported in previous research that taking advantage of correlated and high heritability traits improved

**Table 1. Heritability estimates and corresponding standard errors (in parentheses) for grain yield in five environments.**

Environments	Relationship matrix	
	Genomic ( <i>G</i> )	Pedigree ( <i>A</i> )
Optimal	0.25 (0.09)	0.09 (0.04)
Early heat	0.37 (0.09)	0.32 (0.06)
Late heat	0.46 (0.10)	0.22 (0.06)
Drought	0.47 (0.10)	0.24 (0.06)
Severe drought	0.45 (0.09)	0.41 (0.06)

the prediction accuracy of low-heritability traits in a multivariate prediction model based on simulation (Jia and Jannink, 2012). In our data set, the heritability estimates for grain yield were only based on one replicate of the data, which resulted in a lower heritability estimate for grain yield in the optimal environment using the pedigree relationship matrix (0.09). Nevertheless, the rest of heritabilities were still moderate from 0.22 to 0.47. In contrast to the heritability of grain yield, the heritabilities of secondary traits were relatively high, which should be beneficial to the predictive ability for grain yield. In this study, the relatively high heritabilities of secondary traits contributed to substantial improvement of predictive ability for grain yield in the multivariate prediction model. For example, among five environments, although secondary traits only had moderate correlation with grain yield in the optimal environment, we observed the greatest improvement of predictive ability for grain yield in this environment (84% in genomic relationship, 101% in pedigree relationship), which should be a result of the higher heritabilities of secondary traits and the lower heritability of grain yield. In addition, even when the heritability estimate of the secondary trait was lower than the primary trait, the predictive ability was still improved as a result of the high correlation between those two traits, for example, CT in late heat environment. However, the amount of improvement was indeed affected by the heritability. Accordingly, both heritabilities and correlations of secondary traits with grain yield played significant roles to improve predictive ability for the primary trait in the multivariate prediction model. This was reflected in Fig. 3, where ISE calculated based on the square root of heritability ratio and correlation of secondary traits and grain yield was associated with the improvement of predictive ability.

### Importance of Secondary Traits in the Prediction Model

Multivariate prediction models take advantage of genetically correlated secondary traits (Guo et al., 2014). Hence, including secondary traits in both training and test populations improved the prediction accuracy for grain yield by efficiently employing strong genetic correlations between secondary traits and grain yield. However, we also found that predictive abilities were not increased when the secondary traits only existed in training population (Fig. 1;

**Table 2. Heritability estimates and corresponding standard errors (in parentheses) for secondary traits in five environments using best linear unbiased predictions of secondary traits from multitrait (MT), random regression (RR), and simple repeatability (SR) models.**

Secondary trait†	Environment	MT		.RR		SR	
		G‡	A§	G	A	G	A
CT	Optimal	0.65 (0.09)	0.51(0.06)	0.63 (0.09)	0.38 (0.06)	0.56 (0.09)	0.37 (0.06)
	Early heat	0.55 (0.10)	0.45 (0.06)	0.47 (0.10)	0.39 (0.06)	0.45 (0.10)	0.38 (0.06)
	Late heat	0.42 (0.10)	0.23 (0.06)	0.40 (0.10)	0.17 (0.06)	0.42 (0.09)	0.19 (0.06)
	Drought	0.56 (0.09)	0.39 (0.06)	0.55 (0.10)	0.32 (0.06)	0.55 (0.09)	0.31 (0.06)
	Severe drought	0.32 (0.09)	0.13 (0.05)	0.43 (0.10)	0.25 (0.06)	0.25 (0.09)	0.14 (0.06)
GNDVI	Optimal	0.76 (0.08)	0.53 (0.05)	0.71 (0.09)	0.49 (0.06)	0.70 (0.09)	0.48 (0.06)
	Early heat	0.63 (0.09)	0.56 (0.05)	0.59 (0.09)	0.50 (0.06)	0.57 (0.09)	0.48 (0.06)
	Late heat	0.58 (0.09)	0.40 (0.06)	0.56 (0.10)	0.38 (0.06)	0.53 (0.09)	0.36 (0.06)
	Drought	0.73 (0.08)	0.54 (0.05)	0.72 (0.08)	0.54 (0.05)	0.70 (0.08)	0.51 (0.06)
	Severe drought	0.68 (0.08)	0.42 (0.05)	0.69 (0.09)	0.43 (0.05)	0.64 (0.09)	0.39 (0.06)
RNDVI	Optimal	0.70 (0.09)	0.51 (0.06)	0.67 (0.09)	0.48 (0.06)	0.65 (0.09)	0.45 (0.06)
	Early heat	0.59 (0.09)	0.51 (0.06)	0.58 (0.09)	0.47 (0.06)	0.55 (0.09)	0.45 (0.06)
	Late heat	0.58 (0.09)	0.43 (0.06)	0.56 (0.09)	0.42 (0.06)	0.57 (0.09)	0.40 (0.06)
	Drought	0.63 (0.09)	0.50 (0.06)	0.63 (0.09)	0.54 (0.06)	0.62 (0.09)	0.47 (0.06)
	Severe drought	0.54 (0.09)	0.45 (0.05)	0.53 (0.09)	0.43 (0.06)	0.50 (0.09)	0.38 (0.06)

† CT, canopy temperature; GNDVI, green normalized difference vegetation index; RNDVI, red normalized difference vegetation index.

‡ G, Genomic relationship matrix.

§ A, Pedigree relationship matrix.

**Table 3. Correlations between secondary traits and grain yield and corresponding standard errors (in parentheses) for genetic correlations.**

Secondary trait†	Environment	MT			RR			SR		
		G‡	A§	P¶	G	A	P	G	A	P
CT	Optimal	-0.69 (0.13)	-0.74 (0.11)	-0.34	-0.77 (0.11)	-0.78(0.11)	-0.35	-0.78 (0.12)	-0.84 (0.10)	-0.35
	Early heat	-0.74 (0.10)	-0.80 (0.07)	-0.52	-0.76 (0.10)	-0.77(0.08)	-0.50	-0.76 (0.10)	-0.81 (0.07)	-0.53
	Late heat	-0.90 (0.07)	-0.88 (0.08)	-0.40	-0.90 (0.07)	-0.89(0.08)	-0.39	-0.89 (0.07)	-0.88 (0.09)	-0.40
	Drought	-0.69 (0.11)	-0.61 (0.12)	-0.38	-0.69 (0.11)	-0.61(0.12)	-0.39	-0.69 (0.11)	-0.56 (0.13)	-0.36
	Severe drought	-0.58 (0.15)	-0.65 (0.14)	-0.34	-0.48 (0.15)	-0.45(0.14)	-0.33	-0.50 (0.18)	-0.36 (0.17)	-0.28
GNDVI	Optimal	0.36 (0.16)	0.49 (0.15)	0.20	0.39 (0.16)	0.49(0.15)	0.20	0.41 (0.16)	0.53 (0.15)	0.20
	Early heat	0.45 (0.14)	0.64 (0.08)	0.43	0.46 (0.14)	0.62(0.09)	0.40	0.48 (0.14)	0.66 (0.09)	0.42
	Late heat	0.52 (0.12)	0.47 (0.13)	0.22	0.49 (0.13)	0.43(0.14)	0.21	0.48 (0.13)	0.42 (0.14)	0.20
	Drought	-0.43 (0.12)	-0.47 (0.11)	-0.39	-0.44 (0.12)	-0.48(0.11)	-0.39	-0.42 (0.13)	-0.46 (0.11)	-0.39
	Severe drought	-0.77 (0.07)	-0.62 (0.08)	-0.59	-0.72 (0.08)	-0.56(0.09)	-0.53	-0.70 (0.09)	-0.56 (0.09)	-0.53
RNDVI	Optimal	0.30 (0.17)	0.42 (0.16)	0.19	0.32 (0.17)	0.36(0.16)	0.17	0.36 (0.17)	0.45 (0.16)	0.18
	Early heat	0.49 (0.13)	0.67 (0.08)	0.45	0.55 (0.13)	0.64(0.09)	0.42	0.59 (0.13)	0.71 (0.08)	0.45
	Late heat	0.64 (0.11)	0.56 (0.12)	0.27	0.61 (0.12)	0.52(0.12)	0.26	0.61 (0.12)	0.53 (0.13)	0.26
	Drought	-0.37 (0.14)	-0.45 (0.11)	-0.40	-0.38 (0.14)	-0.51(0.11)	-0.39	-0.35 (0.14)	-0.47 (0.12)	-0.39
	Severe drought	-0.70 (0.09)	-0.59 (0.08)	-0.54	-0.61 (0.11)	-0.44(0.10)	-0.44	-0.57 (0.12)	-0.46 (0.10)	-0.44

† CT, canopy temperature; GNDVI, green normalized difference vegetation index; RNDVI, red normalized difference vegetation index.

‡ G, Genomic relationship matrix.

§ A, Pedigree relationship matrix.

¶ P, phenotypic correlation.

below MV2), which was also reported by Pszczola et al. (2013) and Rutkoski et al. (2016). Based on previous studies, a multivariate prediction model would be superior to a UV model when the trait of interest has low heritability, missing data, or large differences in genetic and residual correlations compared with the secondary traits (Guo et al., 2014; Pszczola et al., 2013). One possible reason

mentioned in Rutkoski et al. (2016) was the moderate heritabilities of grain yield in this data set resulted in the ineffectiveness of multivariate prediction model when secondary traits only in training population. However, it could be favorable in breeding to include secondary traits to make selections when the primary trait is difficult to measure under circumstances including severe weather

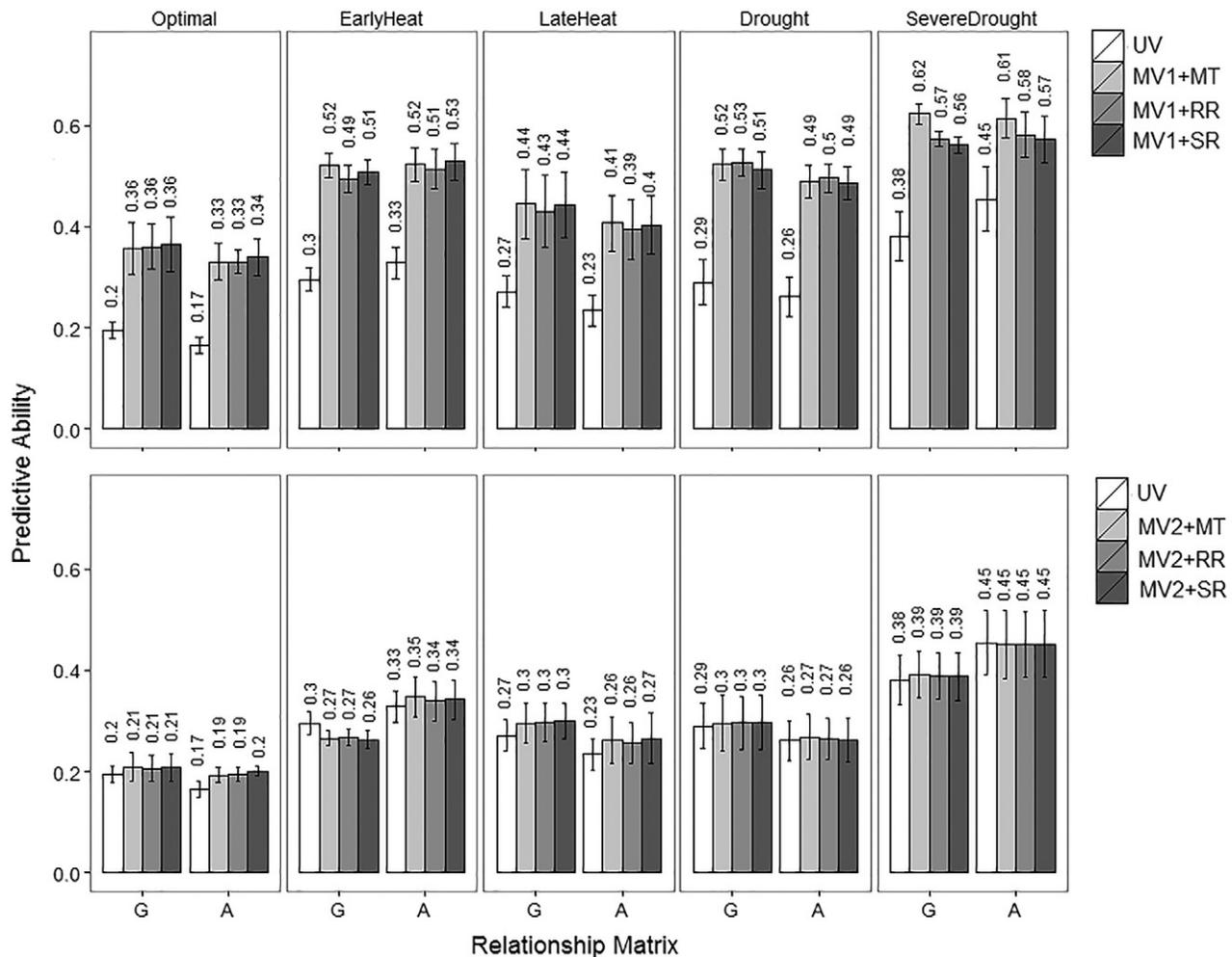


Fig. 1. Predictive ability comparison for grain yield between prediction models with secondary traits (MV1 and MV2) and without secondary trait (UV). MV1, multivariate prediction model with secondary traits in both training and testing populations; MV2, multivariate prediction model with secondary traits in training population only; UV, univariate prediction model with grain yield only; MT/RR/SR, multivariate prediction model MV1 or MV2 using best linear unbiased predictions (BLUPs) of secondary traits from multitrait (MT), random regression (RR), or simple repeatability (SR) model; G/A, genomic/pedigree relationship matrix.

conditions, limited seeds, or expensive cost of phenotyping (Guo et al., 2014; Rutkoski et al., 2016). In this study, the BLUPs for secondary traits and grain yield were predicted separately with data in different replicates (environments). The predictive ability of the UV model was slightly lower when only one replicate was used for grain yield (0.29 on average) than using all three replicates (0.34 on average). However, the predictive ability comparison between the UV model and the multivariate prediction models would be more robust as the environmental covariance between secondary traits and grain yield should be avoided based on data in different environments for them in the multivariate prediction models.

### Single Secondary Trait

Heritability and correlation varied for each secondary trait in each environment, resulting in different predictive abilities for each secondary trait. Canopy temperature predicted grain yield best in the heat-stressed environments, which was consistent with the conclusion of

Mason and Singh (2014), and NDVI predicted grain yield better than CT in the drought-stressed environments. In addition, we observed that including all secondary traits in the pedigree and genomic multivariate prediction model did not necessarily improve the predictive ability for grain yield compared with only a single secondary trait, indicating that the predictive ability in the multivariate prediction model was mostly determined by the secondary trait with the highest ISE. Therefore, a single secondary trait could be selected in different environments, based on the ISE, to provide the best prediction accuracy using a simple BV model. However, if computational resources allowed, incorporating all secondary traits in the prediction model would ensure stable and high prediction accuracy across all environments.

### Genomic versus Pedigree Relationship

Both pedigree and genomic prediction models were compared in this study. Among five environments, the heritability estimates for grain yield from genomic data

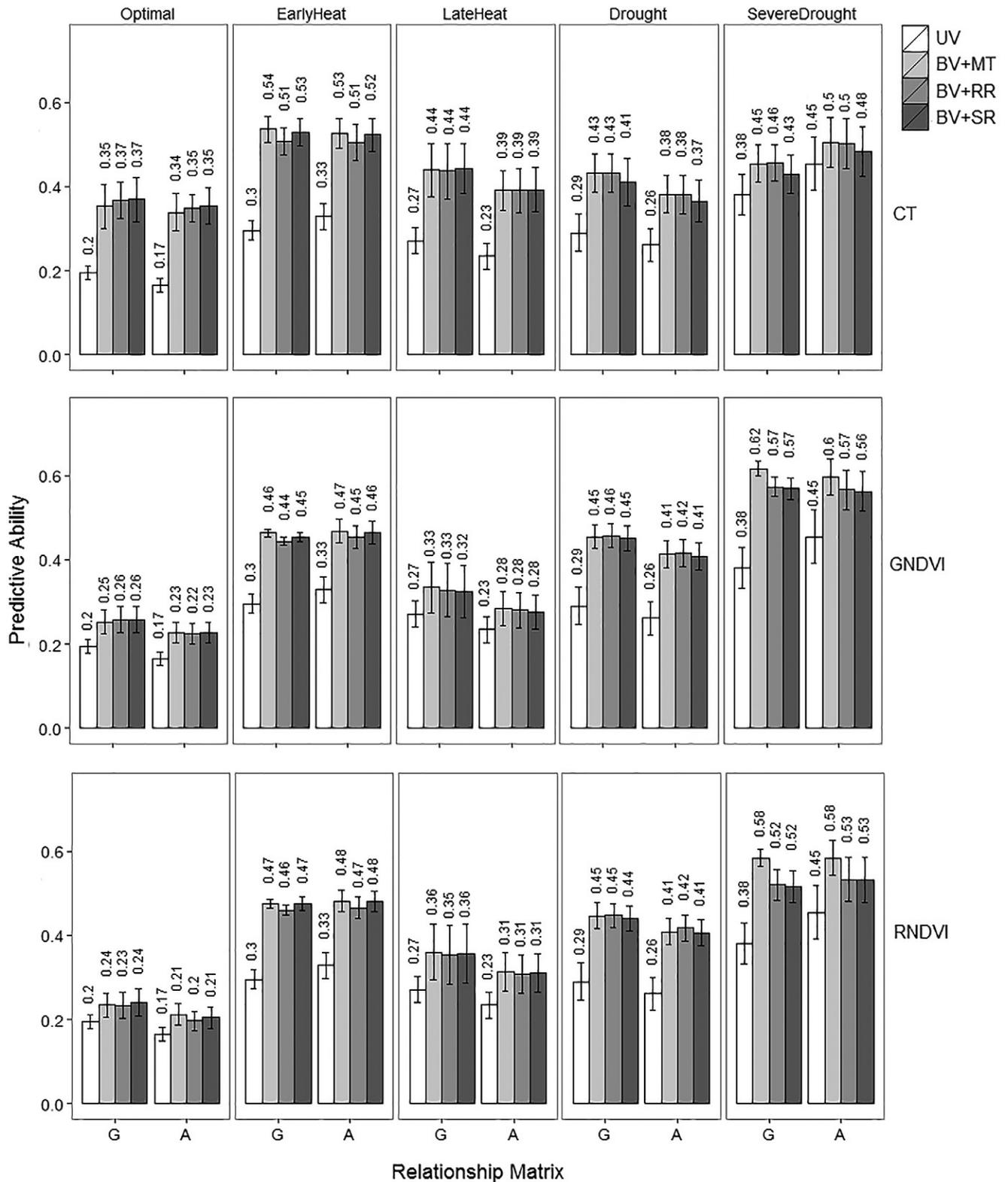


Fig. 2. Predictive ability comparison for grain yield between prediction models with a single secondary trait (BV) and without secondary trait (UV). BV, bivariate prediction model with a single secondary trait in both training and testing populations; UV, univariate prediction model with grain yield only; MT/RR/SR, bivariate prediction model BV using best linear unbiased predictions (BLUPs) of secondary traits from multitrait (MT), random regression (RR), or simple repeatability (SR) model; CT/GNDVI/RNDVI, bivariate prediction model BV including a single secondary trait CT, GNDVI, or RNDVI; G/A, genomic/pedigree relationship matrix.

were higher than those from pedigree based relationship, whereas predictive abilities for grain yield from pedigree

performed better than those from genomic based relationship in the UV model and even in the MV2 model

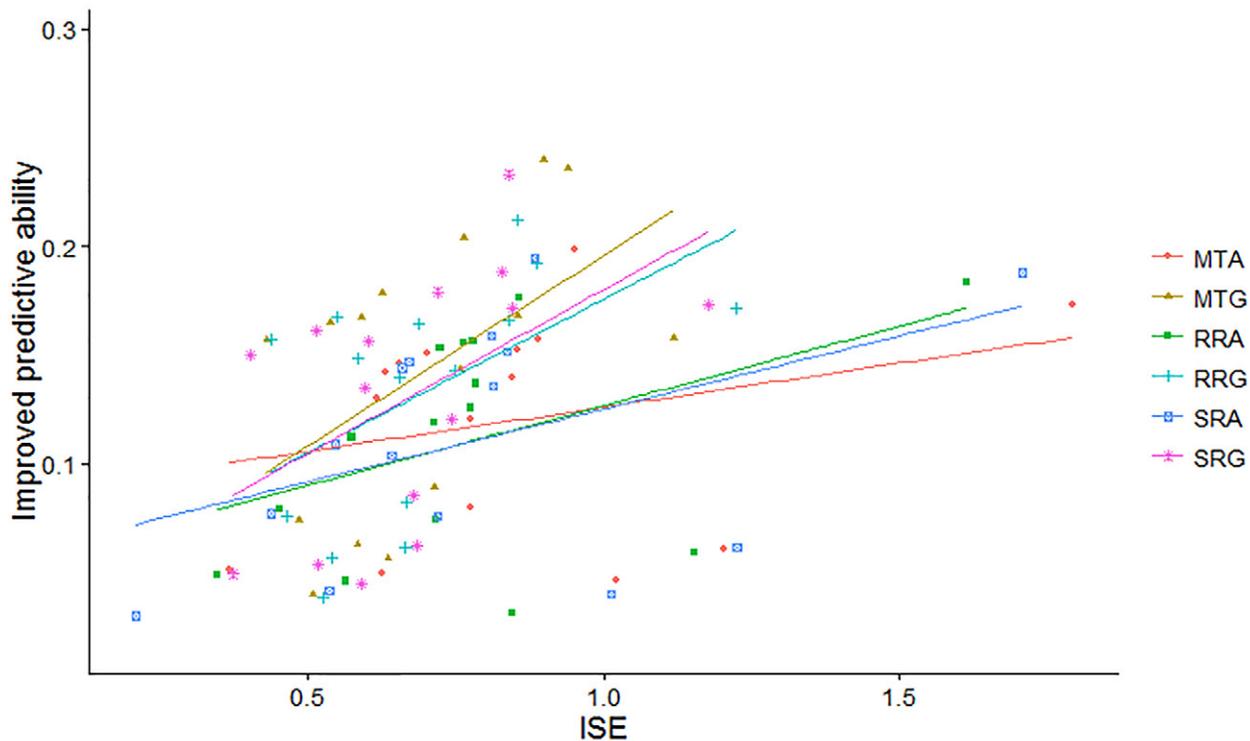


Fig. 3. Relationship between improved predictive ability and indirect selection efficiency (ISE) for single secondary trait. Improved predictive ability, predictive ability for grain yield from the bivariate prediction model (BV) minus predictive ability for grain yield from the univariate prediction model (UV); Model MTA/MTG, multitrait (MT) model based on pedigree (A) or genomic (G) relationship matrix; RRA/RRG, random regression (RR) model based on A or G relationship matrix; SRA/SRG, simple repeatability (SR) model based on A or G relationship matrix.

in early heat and severe drought environments. Furthermore, including secondary traits in the MV1 model also resulted in similar or only slightly higher prediction accuracy from genomic than pedigree based relationship in all environments as was observed by Pszczola et al. (2013). We observed different scales in genomic- and pedigree-based relationship matrices (Supplemental Fig. S2), which could influence the comparability of heritability estimates from two relationships in this data set. In addition, CIMMYT has a reliable and deep pedigree database extending over several generations that contributes to predictive ability. Also, 557 lines were derived from 332 families in this population, giving an average of one to two lines per family. One advantage of genomic relationship vs. pedigree relationship is that it captures the variation within family as a result of Mendelian sampling (Pszczola et al., 2013), and the variance within family in this population may be very limited because of small family sizes compared with the variance among families. However, the relatively higher association between gains in predictive ability and ISE of a single predictive trait using genomic instead of pedigree based relationship indicated the gain in predictive ability from secondary traits was greater when genomic relationships were used.

## Simple Repeatability, Multitrait, and Random Regression Models

### Heritability and Correlation

Best linear unbiased predictions of secondary traits from the MT model provided higher heritability estimates than the other two models, SR and RR, under most environments except severe drought. Similar results were observed in previous studies (Ahrabi, 2015; Iwaisaki et al., 2005; Menéndez-Buxadera et al., 2008; Oh and See, 2008). In contrast to heritability estimates, the correlations between grain yield and secondary traits from the MT model were lower than the other two models under most environments except severe drought. Opposite trends for heritability and correlation were observed by comparing SR, MT, and RR models. Relatively higher heritability estimates of secondary traits among the three models were coupled with relatively lower correlations with grain yield.

### Predictive Abilities

In this study, we observed that predictive abilities for grain yield using BLUPs of secondary traits from the RR model was only slightly better than the SR and MT models in the drought environment. However, in other studies, the RR model was considered to be superior to the SR and MT models. The number of time points in this data set might influence the growth curve adjustment of the RR model, particularly for those regions where data

changes rapidly (Kranis et al., 2007; Mota et al., 2013). Increasing number of the time points will cause over-parameterized analysis in the MT model (Oh and See, 2008). Thus, the small differences between the MT and RR models might be caused by the few time points available in our data set. As for the SR model, it also showed accuracy similar to the other two models. The correlations between BLUPs of secondary traits from three models were close to one, particularly for those between the SR and RR models. This was also reported in a cattle study using those three models (Gernand and König, 2014). The repeatability of secondary traits was high and similar across time points in our data set, which could be another reason for these results. Although the SR model is a simple and fast approach, it may not always be desirable to apply the model in longitudinal data because of limiting assumptions (Kranis et al., 2007). Since the RR model is attractive for longitudinal data because of the advantages of handling many unequally spaced time points and capturing the covariance between or at each time point (Huisman et al., 2002; Meyer, 2005), it would be valuable to investigate this model further using other data sets. Several directions could be taken for the RR model in the future: (i) evaluate the RR model using a data set including more time points to check if it would better model the genetic variation over time (Kranis et al., 2007); (ii) explore the number and location of knots for the spline RR model to improve the accuracy of the growth curve fit (Speidel, 2011); (iii) instead of using information from all time points, analyze estimated breeding values from each time point separately to select the optimum period in genomic prediction for grain yield; and (iv) incorporate genotype  $\times$  environment effects in the prediction model to capture the relationship among environments.

## Conclusion

The predictive abilities for grain yield were improved ~70%, on average, by including BLUPs of secondary traits from the SR, MT, or RR models in the multivariate pedigree and genomic selection models. In this data set, our results showed that the RR or MT model was superior to the SR model in drought-stressed environments, and using a simple SR model for longitudinal data does not negatively affect the predictive ability for grain yield in MV models. However, more studies are necessary to further evaluate these models in other data sets. The advantages of the RR model over the SR and MT models might be more apparent for other data sets from HTP platforms that have more data collection time points within vegetative and grain filling stages. In addition, we observed that the predictive ability would only be improved if both training and test populations contain secondary traits. Single secondary trait could be selected to fit a simple BV model based on the highest ISE for a secondary trait, and this had similar prediction accuracy as when using all secondary traits in a multivariate prediction model.

## Acknowledgments

Program activities were funded by the United States Agency for International Development (USAID) “Feed the Future Initiative” (Cooperative Agreement #AID-OAA-A-13-00051) and by participating US and host country institutions. Partial funding was provided by Hatch project 149-430. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of Cornell University, Kansas State University, CIMMYT, ICAR, and BISA and do not necessarily reflect the view of USAID.

## References

- Ahrabi, M.P. 2015. Heritabilities and genetic correlations for egg weight traits in Iranian fowl by multi trait and random regression models. *Int. J. Adv. Biol. Biom. Res.* 3:108–111.
- Araus, J.L., and J.E. Cairns. 2014. Field high-throughput phenotyping: The new crop breeding frontier. *Trends Plant Sci.* 19:52–61. doi:10.1016/j.tplants.2013.09.008
- Boligon, A., M. Mercadante, R. Lôbo, F. Baldi, and L.G. Albuquerque. 2012. Random regression analyses using B-spline functions to model growth of Nelore cattle. *Animal* 6:212–220. doi:10.1017/S1751731111001534
- Butler, D., B. Cullis, A. Gilmour, and B. Gogel. 2009. Mixed models for S language environments: ASReml-R reference manual. Queensland Dep. of Primary Industries, Queensland, Australia. <https://www.vsnl.co.uk/downloads/asreml/release3/asreml-R.pdf>
- Endelman, J.B. 2011. Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome* 4:250–255. doi:10.3835/plantgenome2011.08.0024
- Endelman, J.B., and J.L. Jannink. 2012. Shrinkage estimation of the realized relationship matrix. *G3: Genes, Genomes, Genet.* 2:1405–1413. doi:10.1534/g3.112.004259
- Falconer, D.S., and T.F. Mackay. 1996. *Introduction to quantitative genetics.* Longman, Harlow, UK.
- Gernand, E., and S. König. 2014. Random regression test-day model for clinical mastitis: Genetic parameters, model comparison, and correlations with indicator traits. *J. Dairy Sci.* 97:3953–3963. doi:10.3168/jds.2013-7830
- Gitelson, A.A., Y.J. Kaufman, and M.N. Merzlyak. 1996. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sens. Environ.* 58:289–298. doi:10.1016/S0034-4257(96)00072-7
- Green, P.J., and B.W. Silverman. 1994. *Nonparametric regression and generalized linear models: A roughness penalty approach.* Chapman & Hall, London, UK. doi:10.1007/978-1-4899-4473-3
- Guo, G., F. Zhao, Y. Wang, Y. Zhang, L. Du, and G. Su. 2014. Comparison of single-trait and multiple-trait genomic prediction models. *BMC Genet.* 15:1. doi:10.1186/1471-2156-15-30
- Huisman, A., R. Veerkamp, and J. Van Arendonk. 2002. Genetic parameters for various random regression models to describe the weight data of pigs. *J. Anim. Sci.* 80:575–582. doi:10.2527/2002.803575x
- Iwaisaki, H., S. Tsuruta, I. Misztal, and J. Bertrand. 2005. Genetic parameters estimated with multitrait and linear spline-random regression models using Gelbvieh early growth data. *J. Anim. Sci.* 83:757–763. doi:10.2527/2005.834757x
- Jia, Y., and J.L. Jannink. 2012. Multiple-trait genomic selection methods increase genetic value prediction accuracy. *Genetics* 192:1513–1522. doi:10.1534/genetics.112.144246
- Kranis, A., G. Su, D. Sorensen, and J. Woolliams. 2007. The application of random regression models in the genetic analysis of monthly egg production in turkeys and a comparison with alternative longitudinal models. *Poult. Sci.* 86:470–475.
- 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. doi:10.1080/01431160110107653
- 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. doi:10.3390/agronomy4020191
- Menéndez-Buxadera, A., C. Carleos, J. Baro, A. Villa, and J. Cañón. 2008. Multi-trait and random regression approaches for addressing the wide range of weaning ages in Asturiana de los Valles beef cattle for genetic parameter estimation. *J. Anim. Sci.* 86:278–286. doi:10.2527/jas.2007-0252

- Meyer, K. 1998. Estimating covariance functions for longitudinal data using a random regression model. *Genet. Sel. Evol.* 30:221–240. doi:10.1186/1297-9686-30-3-221
- Meyer, K. 2000. Random regressions to model phenotypic variation in monthly weights of Australian beef cows. *Livest. Prod. Sci.* 65:19–38. doi:10.1016/S0301-6226(99)00183-9
- Meyer, K. 2005. Random regression analyses using B-splines to model growth of Australian Angus cattle. *Genet. Sel. Evol.* 37:473–500. doi:10.1186/1297-9686-37-6-473
- Meyer, K., and W.G. Hill. 1997. Estimation of genetic and phenotypic covariance functions for longitudinal or ‘repeated’ records by restricted maximum likelihood. *Livest. Prod. Sci.* 47:185–200. doi:10.1016/S0301-6226(96)01414-5
- Miształ, I. 2006. Properties of random regression models using linear splines. *J. Anim. Breed. Genet.* 123:74–80. doi:10.1111/j.1439-0388.2006.00582.x
- Mota, R., L. Marques, P. Lopes, L. Da Silva, F. Neto, M. de Resende, and R. Torres. 2013. Genetic evaluation using multi-trait and random regression models in Simmental beef cattle. *Genet. Mol. Res.* 12:2465–2480. doi:10.4238/2013.July.24.2
- Mrode, R.A. 2005. Linear models for the prediction of animal breeding values. CABI Publishing, London, UK. doi:10.1079/9780851990002.0000
- Narjesi, V., M. Mardi, E.M. Hervas, A. Azadi, M.R. Naghavi, M. Ebrahimi, and A.A. Zali. 2015. Analysis of quantitative trait loci (QTL) for grain yield and agronomic traits in wheat (*Triticum aestivum* L.) under normal and salt-stress conditions. *Plant Mol. Biol. Report.* 33:2030–2040. doi:10.1007/s11105-015-0876-8
- Oh, S.H., and M. See. 2008. Comparison of genetic parameter estimates of total sperm cells of boars between random regression and multiple trait animal models. *Asian-Aust. J. Anim. Sci.* 21:923–927.
- Poland, J., J. Endelman, J. Dawson, J. Rutkoski, S. Wu, Y. Manes, S. Dreisigacker, J. Crossa, H. Sánchez-Villeda, M. Sorrells, and J.L. Jannink. 2012. Genomic selection in wheat breeding using genotyping-by-sequencing. *Plant Genome* 5:103–113. doi:10.3835/plantgenome2012.06.0006
- Pszczola, M., R. Veerkamp, Y. De Haas, E. Wall, T. Strabel, and M. Calus. 2013. Effect of predictor traits on accuracy of genomic breeding values for feed intake based on a limited cow reference population. *Animal* 7:1759–1768. doi:10.1017/S175173111300150X
- Quarmby, N., M. Milnes, T. Hindle, and N. Silleos. 1993. The use of multi-temporal NDVI measurements from AVHRR data for crop yield estimation and prediction. *Int. J. Remote Sens.* 14:199–210. doi:10.1080/01431169308904332
- R Development Core Team. 2010. R: A language and environment for statistical computing. R Foundation for Stat. Comput., Vienna, Austria.
- Rees, D., K. Sayre, E. Acevedo, T. Nava Sanchez, Z. Lu, E. Zeiger, and A. Limon. 1993. Canopy temperatures of wheat: Relationships with yield and potential as a technique for early generation selection. CIMMYT Wheat Special Report (WPSR). CIMMYT, Mexico, DF.
- Rutkoski, J., J. Benson, Y. Jia, G. Brown-Guedira, J.L. Jannink, and M. Sorrells. 2012. Evaluation of genomic prediction methods for Fusarium head blight resistance in wheat. *Plant Genome* 5:51–61. doi:10.3835/plantgenome2012.02.0001
- Rutkoski, J.E., E.L. Heffner, and M.E. Sorrells. 2011. Genomic selection for durable stem rust resistance in wheat. *Euphytica* 179:161–173. doi:10.1007/s10681-010-0301-1
- Rutkoski, J., J. Poland, S. Mondal, E. Autrique, L.G. Pérez, J. Crossa, M. Reynolds, and R. Singh. 2016. Canopy temperature and vegetation indices from high-throughput phenotyping improve accuracy of pedigree and genomic selection for grain yield in wheat. *G3: Genes, Genomes, Genet.* 6:2799–2808. doi:10.1534/g3.116.032888
- Speidel, S.E. 2011. Random regression models for the prediction of days to finish in beef cattle. Ph.D. diss. Colorado State Univ., Ft. Collins.
- Tucker, C.J. 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* 8:127–150. doi:10.1016/0034-4257(79)90013-0
- White, I., R. Thompson, and S. Brotherstone. 1999. Genetic and environmental smoothing of lactation curves with cubic splines. *J. Dairy Sci.* 82:632–638. doi:10.3168/jds.S0022-0302(99)75277-X
- Yang, W., Z. Guo, C. Huang, L. Duan, G. Chen, N. Jiang, W. Fang, H. Feng, W. Xie, X. Lian, G. Wang, Q. Luo, Q. Zhang, Q. Liu, and L. Xiong. 2014. Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. *Nat. Commun.* 5:5087. doi:10.1038/ncomms6087