

**GENOMIC SELECTION IN A SINGLE CROSS
DOUBLED-HAPLOID WHEAT POPULATION**

by

Jiayin (Susan) Song

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Forestry)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

February 2017

© Jiayin (Susan) Song, 2017

Abstract

A traditional wheat breeding program normally takes 7 to 12 years to develop a new cultivar to be eligible for commercial release. Genomic selection (GS), which uses single-nucleotide polymorphism (SNP) marker information to predict breeding values, has been proven to be an efficient method to accelerate the lengthy breeding process and increase the resultant gain in many animal and plant species. In this study, two GS algorithms, Genomic Best Linear Unbiased Prediction (GBLUP) and Reproducing Kernel Hilbert Space (RKHS) regression, were evaluated using grain yield data generated from a single hard red winter wheat (*Triticum aestivum* L.) full-sib doubled-haploid (DH) population in two consecutive generations. In each generation, a total of 257 individuals were genotyped with 14,028 SNP markers using “Genotyping-by-Sequencing” (GBS). Due to the uniformity of genetic material across generations, year effect was considered as an environmental factor or replication for the analysis. Potential upward bias in model’s predictive accuracy was estimated by comparing the within-year cross-validation scheme with the cross-year prediction scheme. The effect of SNP marker number on the models’ predictive ability was also analyzed by creating SNP subsets filtered with absolute pairwise correlation (t) value. In general, RKHS produced higher predictive ability than GBLUP for predicting grain yield in this population. A 32 and 38% decrease in predictive ability was observed for GBLUP and RKHS models, respectively, when comparing within-year cross-validation and cross-year prediction models’ results. A t value of 0.4 could produce a similar predictive ability compared to using the unfiltered full SNP set, providing less computation- and time-consuming strategy. In the context of an ongoing breeding program, this study also demonstrated confidence of line selection based on GS results, advocating the implementation of GS in wheat variety development.

Preface

This research is a part of “Establishing Translational Genomics for Oklahoma Wheat Improvement” project designed by Brett Carver and Charles Chen; sequencing data configuration was carried out by Shuzhen Sun and Charles Chen; and the DNA extraction was done by Carol Power and Karyn Willyerd. All the people mentioned above are affiliated with Oklahoma State University.

Table of Contents

Abstract.....	ii
Preface.....	iii
Table of Contents	iv
List of Tables	vi
List of Figures.....	vii
List of Abbreviations	ix
Acknowledgements	x
Dedication	xi
Introduction.....	1
Materials and methods	4
Phenotypic data.....	4
Genotypic data	5
Statistical models	5
Within- and cross-year prediction.....	7
Marker selection.....	8
Consistency of elite line selection.....	9
Results	10
Data description	10
Genomic prediction model performance	11
Effects of missing genotype data and imputation methods	11
GS predictability evaluation	11

Within-year prediction using 2014 as training population	12
Within-year prediction using 2015 as training population	12
Cross-year prediction using 2014 as training population	13
Cross-year prediction using 2015 as training population	17
Marker selection.....	20
Consistency of elite line selection.....	23
Discussion.....	25
Model comparison	25
Missing data imputation.....	29
Marker selection.....	29
Conclusion	31
Bibliography	32
Appendices.....	38
Appendix A Effect of bandwidth parameter h on model predictive ability across training population, model composition, and number of markers employed.....	38

List of Tables

Table 1 Summary of infection assessment statistics (mean and standard deviation)	10
Table 2 Total number of SNPs in each subgroup	10
Table 3 Best-performing models and the number of SNPs required (SE = standard error).	21
Table 4 Number of SNP markers within each correlation group based on whole population data	22

List of Figures

Figure 1 Comparison of two missing data imputation methods, EM and Mean, based on the predictive ability from GBLUP cross-validation model (with SNP effect only) across a gradient of missing marker data ratio (Missingness); TP = training population.	14
Figure 2 Comparison of two missing data imputation methods, EM and Mean, based on the predictive ability from RKHS cross-validation model (with SNP effect only) across a gradient of missing marker data ratio (Missingness); TP = training population; bandwidth parameter was set to 0.1 for all models.	15
Figure 3 GBLUP and RKHS prediction accuracies from the best-performing models using individuals from year 2014 as training population (TP). Within each model type, within-year cross-validation predictive ability was compared to cross-year predictive ability. G = only marker effect and G+HD = marker effect and heading date as models' covariates.	16
Figure 4 GBLUP and RKHS predictive ability from the best-performing models using individuals from year 2015 as training population (TP). Within each model type, within-year cross-validation predictive ability was compared to cross-year predictive ability. G = only marker effect; G+HD = marker effect and heading date as covariate; G+Rust = marker effect and disease index as covariate; G+HD+Rust = all three components in the model.	19
Figure 5 Predictive ability from the best within-year cross-validation model (Within: year 2015 RKHS model with marker effect and both heading date and disease index as covariates), and the best cross year prediction model (Cross: year 2014 predicting 2015 RKHS model with marker effect and heading date as covariate) across subsets of marker filtered by absolute pairwise correlation threshold t	22

Figure 6 Trend of average ranking distance over the number of individuals considered. Analysis started with the ranking distance of the best-performing individual in the year 2014, proceeded by adding the next best individual's ranking distance and taking the mean, until all 257 individuals in the population were included	24
Figure 7 Comparison of bandwidth parameter h based on predictive ability from RKHS (with year 2014 and 2015 as training population respectively) cross validation model (with SNP effect only) across marker missingness levels	38
Figure 8 Comparison of bandwidth parameter h based on predictive ability from RKHS cross year prediction model (with year 2014 and 2015 as training population respectively) across different model components: G =only marker effect in the model; $G+HD(cor)$ =marker effect in the model with heading date-corrected phenotype as response variable; $G+HD(cov)$ =marker effect and heading date as covariate in the model; $G+Rust(cov)$ =marker effect and disease index as covariate in the model; $G+HD+Rust(cov)$ =marker effect and both heading date and disease index as covariates in the model	39

List of Abbreviations

DH – Doubled-Haploid

EM – Expectation Maximization

GBLUP – Genomic Best Linear Unbiased Prediction

GBS – Genotyping-By-Sequencing

GEBV – Genomic Estimated Breeding Value

GS – Genomic Selection

HD – Days to Heading

LD – Linkage Disequilibrium

Obs – Observed Phenotype

RI – Rust Incidence

RKHS – Reproducing Kernel Hilbert Space

RS – Rust Severity

SNP – Single-Nucleotide Polymorphism

TBV – True Breeding Value

TP – Training Population

Acknowledgements

I would like to express my earnest gratitude to my supervisor, Dr. Yousry A El-Kassaby, for his great mentorship in research as well as in life throughout this journey. I truly appreciate all the guidance, and invaluable opportunities he provided me, without which completion of this thesis would not have been possible. Heartfelt thanks to all my committee members: Dr. Charles Chen, for introducing me to the world of crop science, and supporting me the entire time with his expertise in wheat breeding and bioinformatics; Dr. Jaroslav Klápště, for countless hours of teaching in statistics and trouble-shooting with my data analysis, as well as his patience with all my questions; Dr. Richard Hamelin, for his unique insight into the research project and warm encouragements that motivates me through this program.

In addition, I would like to thank everyone in our lab, Blaise Ratcliffe, Omnia Gamal El-Dien, Frances Thistlethwaite, Yang Liu, Faisal Al-harbi, and Qing Wang, for creating such a relaxed and pleasant working environment. I am extremely fortunate to be a part of this talented and diverse group, and I cannot be more grateful for all the help and support I received from this caring family. Last but not least, special thanks to my parents for believing in me from the very start. Their unconditional love and support gives me the momentum to keep moving forward.

To my family and friends

Introduction

As the most widely planted cereal crop around the globe, wheat is one of the world's most important food and protein resources, and has the greatest world trade among all crops (United States Department of Agriculture, 2016). In order to sustain the species' vitality, it is pivotal to frequently update the list of high-yielding cultivars meeting the current breeding goals and adapted to the ever-changing environment.

A traditional wheat breeding program normally requires at least 7 or 12 years for spring and winter wheat, respectively, before a developed cultivar is ready for commercial release (Baenziger & Depauw, 2009). Once the main objective is determined, for example, to improve grain yield and adaptability, the breeding program is initiated with the hybridization stage. In this stage, crosses are made between parental lines with the traits of interest. Based on the specificity of the parents, the most appropriate cross type is carried out to produce the F_1 generation. Following hybridization, efforts are made to reduce among populations' heterozygosity and heterogeneity, while concurrently selecting desirable progenies for further assessment. Since wheat is obligatory selfer, the proportion of heterozygous loci decreases by 50% after each generation. As the breeding program advances until reaching an acceptable level of homozygosity, a considerable amount of alleles contributing to the target traits could be lost. With this in mind, retaining high level of desired alleles in the population is a difficult task, yet it is a crucial matter for breeding practices. When the level of variability within a population is reduced to a manageable level, a round of selection is made to create new elite lines. The main objective of such breeding and selection programs is traits evaluation, including the physical characteristics of grain, lines endurance under biotic and abiotic stresses, as well as resistance to a suite of diseases. Before elite lines commercial release, the finalists from the selection stage must go through extensive evaluation in replicated

yield trials. Due to the highly correlated environmental effects within a specific year, three years of replicated yield evaluation is commonly practiced and this is considered as the minimal evaluation for variety release. Finally, after 7 to 12 years of hybridization, evaluation, and selection, often only a single cultivar is released.

As the demand for wheat consumption is exceeding the current supply (United States Department of Agriculture, 2016b), an estimated 1.6% annual minimum increase in wheat production is required to fulfill the projected demand in 2020 of 760 million tons (Tadesse et al., 2016). Looking further ahead, Alexandratos (2009) predicted the world population to exceed nine billion by 2050, driving the demand to reach 900 million tons. Given the present average increase rate of 1.1% (Tadesse et al., 2016), the mismatch between the projected supply and demand is an obvious global challenge. As a result, it is imperative to incorporate emerging technologies into wheat breeding programs to ensure productivity meeting these challenges.

Genomic selection (GS), which employs single-nucleotide polymorphism (SNP) markers across the entire genome to predict an individual's performance in quantitative traits (Meuwissen, Hayes, and Goddard, 2001), is a proven method that can accelerate breeding process. Genomic selection has become a common practice in animal and plant breeding (Heslot, Jannink, & Sorrells, 2015; Bassi et al. 2016). The satisfying outcome surely inspired scientists in the field of agriculture; for example, Bernardo & Yu (2007) carried out a simulation study demonstrating the advantage of GS in comparison to marker-assisted selection in maize. Two years later, de los Campos et al. (2009) were the first to incorporate GS in wheat breeding by confirming that the inclusion of SNP markers resulted in improvement of GS model's performance in predicting average grain yield. Since then, GS has gained increased acceptance in wheat breeding studies. Initial applications were focused on exploration of only additive genetic variation among individuals, until 2010 when de los

Campos et al. (2009) extended the work of Gianola & Kaam (2008) and utilized the Reproducing Kernel Hilbert Spaces (RKHS) method to account for epistatic effect in addition to additive effect, and evaluated the method's potential in wheat breeding. Subsequently, wheat breeders conducted several studies comparing the GBLUP model with the RKHS (Crossa et al., 2010; 2011; He et al., 2016; Huang et al., 2016), without any conclusive advantage to either method. Ultimately, while traditional phenotypic selection is considered time-, resources-, and space-consuming, the use of GS with its genotypic information, made it possible to predict adult plants' performance from information generated at the early seedlings stage, creating a paradigm shift where genomic information can be used to predict phenotypes to substitute the phenotype-dependent field evaluation, thus the effort and investment on field assessment for phenotypes are substantially reduced (Baenziger & Depauw, 2009). In addition, GS could contribute to replicated yield testing by considering the genotype by environment (G x E) effects in the prediction model and then select the best lines across several environments.

It has been a decade since the first study of GS in plant breeding was published (Bernardo & Yu, 2007). Substantial amount of evidence has shown the potential of GS in wheat, with most of them focusing on increasing models' predictive accuracies. However, very few studies have considered the practical implications of these results in the context of a breeding program. To bring forth GS's practicality, the objectives of the current study aimed to address: 1) GS algorithms performance in wheat grain yield across environments (represented by generations due to the uniformity of genotypes in both generations) in the line evaluation stage; 2) investigate the possible upward bias in predictive ability from within-year cross-validation compared to cross-year prediction; and 3) evaluate the effect of SNP marker information on GS predictive ability.

Materials and methods

Phenotypic data

The efficiency of genomic selection across successive generations was evaluated using a doubled haploid (DH) population derived from a single cross between two hard red winter wheat (*Triticum aestivum* L.) cultivars, ‘Duster’ (Edwards et al., 2012) and ‘Billings’ (Hunger et al., 2014). In total, 282 DH lines were developed, among which 257 lines were evaluated for grain yield (bushels per acre) for 2014 and 2015 at the Agronomy Research Station in Stillwater, Oklahoma. In both year 2014 and 2015, among these 257 lines, 239 lines were replicated twice, and the remaining 18 were screened for grain yield only once. When applicable, the mean of the two replications was taken to represent individual line’s phenotype. Since the genotypes evaluated were identical for 2014 and 2015, the year effect was considered as environmental replication during data analysis. Days to heading (HD) were recorded for every individual plant for both 2014 and 2015, as the variability of HD (measured as the duration from planting to heading) in wheat is indicative of adaptability to its growing environment (Kiseleva et al., 2016). In addition, stripe rust (*Puccinia striiformis*) disease infection was observed in 2015’s field trial; the severity of strip rust was also recorded. Assessment of the rust infection was carried out in between May 5th and May 11th; at each time points, the severity (RS₅, RS₁₁) and incidence (RI₅, RI₁₁) of the disease were recorded for each individual line. Severity of infection was rated between 0 and 5, with 0 being no infection and 5 being the highest level of infection. Incidence of infection was estimated by the percentage of affected area within each plot, ranging from 0 to 100%.

Genotypic data

A next-generation sequencing technology, namely Genotyping-By-Sequencing (GBS, Elshire et al. 2011), was employed to generate genotypic data. The details of enzyme selection, library construction, and SNP data analysis could be found in Poland et al. (2012) and Li et al. (2015). In total, 14,028 SNP markers were generated for these Duster x Billings DH lines, prior to other data treatment like filtering and imputation.

To investigate the impact of missing data on predictive ability, the SNP markers were grouped into 5 subsets based on the call rate of 0.25, 0.4, 0.5, 0.6, and 0.75. Two imputation methods were employed to interpret missing data: mean imputation which uses average genotypic value of each SNP locus for all missing data, and the expectation maximization (EM) algorithm (Poland et al. 2012).

Statistical models

The performance of genomic prediction on grain yield of the 257 DH lines was evaluated by two different types of model: 1) Genomic Best Linear Unbiased Prediction (GBLUP), a parametric model which utilized to only account for additive genetic effect, and 2) Reproducing Kernel Hilbert Space (RKHS) regression, a semi-parametric model that also considers correlations between markers, or epistatic effect.

Let n be the number of genotypes, and m be the number of SNP markers. GBLUP model takes the form:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{g} + \boldsymbol{\varepsilon} \quad (1)$$

where \mathbf{y} is a vector of phenotypic records of dimension $n \times 1$, $\boldsymbol{\beta}$ is a vector of fixed effects containing the common intercept and other terms such as heading date and rust infection status,

and \mathbf{X} is its corresponding design matrix, \mathbf{g} is an $\mathbf{n} \times \mathbf{1}$ vector of genomic breeding values, which were assumed to follow a normal distribution $\mathbf{g} \sim N(\mathbf{0}, \mathbf{G}\sigma_A^2)$, for which σ_A^2 is the additive genetic variance, and \mathbf{G} is the realized relationship matrix constructed following VanRaden (2008):

$$\mathbf{G} = \frac{\mathbf{Z}\mathbf{Z}'}{2\sum \mathbf{p}_i(1 - \mathbf{p}_i)} \quad (2)$$

where \mathbf{Z} is an $\mathbf{n} \times \mathbf{m}$ matrix whose elements are defined as $\mathbf{a} - 2(\mathbf{p}_i - 0.5)$ with \mathbf{a} denoting the marker value as -1 (homozygote), 0 (heterozygote), and 1 (other homozygote), and \mathbf{p}_i is the frequency of the second allele at locus i , $\boldsymbol{\varepsilon}$ is an $\mathbf{n} \times \mathbf{1}$ vector of residuals with $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{I}\sigma_\varepsilon^2)$, with \mathbf{I} denoting an identity matrix of order \mathbf{n} , and σ_ε^2 denoting the residual variance. The GBLUP model was implemented using the R package “rrBLUP” (Endelman, 2011).

The semi-parametric RKHS model was carried out using a single Gaussian kernel. It is represented as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} + \boldsymbol{\varepsilon} \quad (3)$$

which has a similar form to GBLUP, but with a different assumption of $\mathbf{u} \sim N(\mathbf{0}, \mathbf{K}\sigma_u^2)$, where \mathbf{K} is an positive definite kernel matrix of dimension $\mathbf{n} \times \mathbf{n}$, whose elements were the average squared-Euclidean distance between genotypes evaluated using the Gaussian kernel:

$$K(\mathbf{x}_j, \mathbf{x}_k) = \exp[-h \times \frac{\sum_{l=1}^m (\mathbf{x}_{jl} - \mathbf{x}_{kl})^2}{m}] \quad (4)$$

where \mathbf{x}_{jl} denotes the value of l^{th} marker of individual j , h is the bandwidth parameter that determines the speed of decay of marker correlation as they get further apart in space. The RKHS model was implemented using the Bayesian approach in the R package “BGLR” (de los Campos et al., 2010).

The variables used to describe days to heading (HD) and rust infection (RS and RI), were incorporated into the genomic prediction models using two methods: either directly as covariates with fixed effects, or as correctors for the response variable, i.e. the phenotypes. The correction step was carried out using a simple linear model with the observed phenotypes as the response variable, and one of the HD, RS or RI as explanatory variable. Residuals from the models were obtained to serve as the response variable in the genomic prediction models as “corrected phenotypes”. We implemented these two methods for fixed-effect variables in order to explore the differences in model behavior.

Within- and cross-year prediction

Within-year cross-validation was performed for both 2014 and 2015 field evaluations separately. For each year, the data was randomly divided into ten folds, with nine folds as training set and one fold as validation set. Each run was repeated five times with different folding. Evaluation of model predictive ability was based on the Pearson product-moment correlation between the genomic estimated breeding value (GEBV) and the observed phenotype (Obs) value $r_{GS} = r_{GEBV,Obs}$ of individuals in the validation set. First, only the random marker effect was included in the genomic prediction models, then HD was added into the models either as a fixed covariate or as a phenotype corrector. In 2015, RS and RI initially underwent a pre-selection, where each of the variables was introduced into the same prediction model with only random marker effect. These models were assessed based on their predictive abilities to choose the best variable to represent rust infection, which later on was included in the cross-validation model in the same fashion as HD.

In the case of cross-year prediction, each of the two years’ data was treated as the training population, which followed the same scheme with ten folds and five replicates as the within-year

cross-validation to obtain the GEBVs, to predict the performance in the other year. Since all genotypes remained the same between the field evaluations in two years, different years were considered as different environmental replications, and the predictive ability was estimated by the Pearson product-moment correlation between the GEBV obtained from the training cycle and the observed phenotype value in the predicted breeding cycle.

For each scenario, both GBLUP and RKHS models were implemented for evaluation of predictability. Model assessment was conducted across the combinations of five gradients of marker missing data ratio and two imputation methods. A grid search was also carried out for the bandwidth parameter h after the optimal composition of the RKHS model was acquired under each scheme.

Marker selection

To discuss the redundancy that might have been caused by the correlation and linkage of SNP markers, we examined the marker information most efficiently exploited for prediction. The evaluation of marker selection also followed a ten-fold cross-validation scheme. Starting with the marker matrix at the missing data ratio determined by the models having the highest predictive ability from within-year cross-validation and cross-year prediction, a matrix of marker pairwise correlation was then calculated using the R package “corpcor” (Opgen-Rhein & Strimmer, 2007; Schäfer & Strimmer, 2005). This matrix construction, along with the succeeding steps, was carried out within each fold using the marker information from the training population (TP) only. The set of markers were subsequently filtered by removing the ones with any absolute pairwise correlation higher than a threshold value t ($t = 1.0, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2$). The resulting SNPs

were then used to construct the genomic relationship matrix for fitting the two best genomic prediction models in order to obtain predictive ability.

Consistency of elite line selection

For both years, the individuals were ranked based on their GEBVs from the best-performing models. There were four models within each year: best cross-validation models using GBLUP and RKHS, and the best cross-year prediction models using GBLUP and RKHS. Individual with the lowest GEBV was ranked the first, and individual with the highest GEBV was ranked the last (257th). As a result, each individual was assigned four ranking in each year. The mean of these four indexes was taken to represent each individual's ranking within that year. The entire population was then sorted by their mean rank indexes in year 2014 from the highest to the lowest, and the ranking distance (d) was calculated as the absolute difference between the individual's mean rank index in 2014 and its corresponding mean rank index in 2015. Starting with the ranking distance of the highest-GEBV individual in 2014, by adding on the absolute difference of the next individual in the hierarchy (which is the second-highest-GEBV individual), the average ranking distance (\bar{d}) was estimated by the mean of this total, and so on:

$$\bar{d}_n = \frac{\sum_{i=1}^n d_i}{n} \quad (n = 1, 2, \dots, 257) \quad (5)$$

The final product was a vector of length 257, which allowed observing the change in the average ranking distance from the best-performing individual to the worst.

In order to achieve consistent selection of elite lines from this population, a relatively shorter distance for highly-ranked individuals were expected, so that they are more likely to be chosen for further breeding regardless of the environmental differences.

Results

Data description

The average HD in 2014 was 170.37 (± 1.93) days, slightly longer than that of 2015 (161.81 ± 1.13 days) ($p < 0.0001$). The severity and spread of rust infection had the same shift in heading date (Table 1). Grain yield in 2014 had a mean of 23.49 (± 3.96) bushels acre⁻¹, which was lower than the mean grain yield in 2015 (33.13 ± 6.60 bushels acre⁻¹) ($p < 0.0001$). Grain yield in 2014 and 2015 had a positive correlation of 0.417. Before any data treatment, the overall SNP call rate averaged at ~0.6. Filtering for SNP's call rate higher than 0.75, only 4010 GBS SNPs remained; however, 12,944 SNPs can be obtained if considering SNPs call rate higher than 0.25. The total number of SNPs in each missing data subgroup is shown in Table 2.

Table 1 Summary of infection assessment statistics (mean and standard deviation)

Time point	Severity (RS)		Incidence (RI)	
	Mean	SD	Mean	SD
May 5 th	3.08	2.010	8.38	12.193
May 11 th	3.32	1.933	12.68	18.846

Table 2 Total number of SNPs in each subgroup

	SNP call rate threshold				
	0.25	0.4	0.5	0.6	0.75
Total number of SNPs	12,994	9,244	7,260	5,726	4,010

Genomic prediction model performance

Effects of missing genotype data and imputation methods

Two imputation methods, mean and EM, were evaluated by the within-year cross-validation models across a gradient of missing marker data ratio using 2014 and 2015 population, respectively (Figures 1 and 2). In the evaluation of SNPs call rate impact on predictive ability, only random marker effect was included in the tested models. In general, EM imputation outperformed the mean imputation in every scenario; effects of SNPs call rate was little in 2015, as using the 4,010 SNPs that have call rate higher than 0.75 generated similar results with the rest of SNP call rate categories for both imputation methods. However, the effect of SNP call rate was more significant in 2014, where the best predictability was observed when SNPs with call rates > 0.6 were used, and the difference in predictability can be as large as 3% (call rate 0.6 imputed with EM versus call rate 0.25 imputed with Mean, see Figure 1). Also, due to its observable superiority, EM imputation was used for all the following analyses (Figures 1 and 2).

GS predictability evaluation

Given the considerable number of models tested in this study and for the purpose of simplifying their comparisons, models with the highest predictive abilities were chosen to represent the four tested model's scenarios. In each scenario, the best-performing GBLUP and RKHS models were selected for each variable combination: SNP marker only, SNP marker + HD, SNP marker + rust rating, and SNP marker + HD + rust rating. The latter two combinations were only available to models that used 2015 as training population.

Within-year prediction using 2014 as training population

With only the realized relationship matrix \mathbf{G} in the model, GBLUP and RKHS had similar performances when 2014's grain yield was used as the training population for within-year cross-validation: both models yielded the highest predictive ability of ~ 0.57 with the SNP dataset that had SNPs call rate higher than 0.6 and was imputed with EM algorithm (Figure 3). When HD was included in the model either directly as a covariate or indirectly as a corrector for the phenotype, the predictive ability decreased for GBLUP but increased for RKHS. As for predictor variables included in predictive models, both GBLUP and RKHS performed better (i.e., produced higher prediction accuracies). However, GBLUP performed slightly better when the HD was used as a phenotype corrector; RKHS generated a better predictive ability when HD was included as a covariate. Both models required more SNPs to attain their highest predictive abilities (i.e., correlation between GEBV and Obs) (Table 3).

Within-year prediction using 2015 as training population

Models trained with 2015's grain yield data, for within-year cross-validation with only marker effect, RKHS resulted in a higher predictive ability (0.600 ± 0.00412) than GBLUP (0.569 ± 0.00669). Similar with year 2014, the inclusion of HD did not show improvement in model performance for GBLUP, while a modest improvement for RKHS was observed when fitting HD as a covariate. In addition to HD variation, the impact of the rust infection on year 2015 prediction models was evaluated in the same fashion as HD. Both models showed better predictive ability performance with rust infection variables included compared to models that had only HD. GBLUP and RKHS achieved their highest predictive abilities when RS_{11} and RI_5 were included as covariates, respectively. Further, fitting both HD and RI_5 as covariates resulted in the best model

performance for RKHS; interestingly, the GBLUP showed the worst model performance when including both HD and RS_{11} as covariates (Figure 4).

Cross-year prediction using 2014 as training population

Trained with grain yield data in 2014 to predict 2015's grain yield resulted in approximate 42 and 39% reduction in predictive ability for GBLUP and RKHS models, respectively. Using linear GBLUP, models with only SNP marker data produced the best cross-year predictive ability (0.350 ± 0.00215); inclusion of covariates showed a negative impact on the model performance (Figure 3). On the contrary, RKHS prediction performed better when HD was included as a covariate, and this model, in fact, was the best predictive model among all cross-year predictions (0.423 ± 0.00304). Overall, predictive ability for cross-year prediction for 2014 yield data was at $0.354 (\pm 0.037)$, significantly lower than within-year cross-validation (Figure 3).

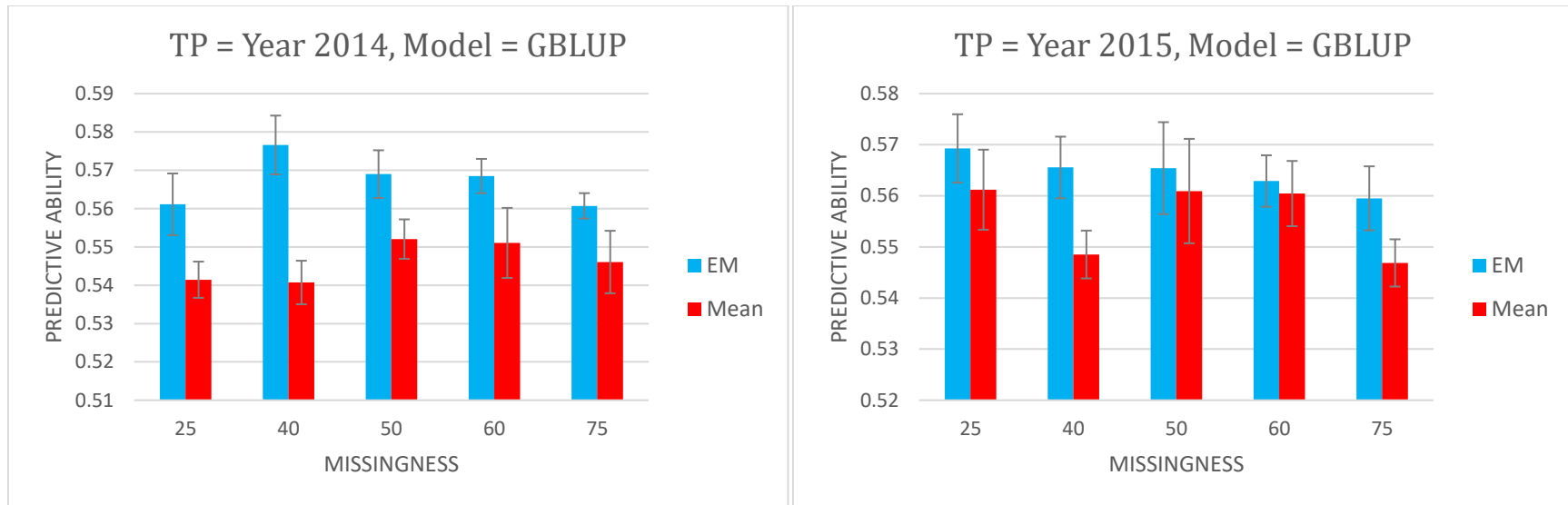


Figure 1 Comparison of two missing data imputation methods, EM and Mean, based on the predictive ability from GBLUP cross-validation model (with SNP effect only) across a gradient of missing marker data ratio (Missingness); TP = training population.

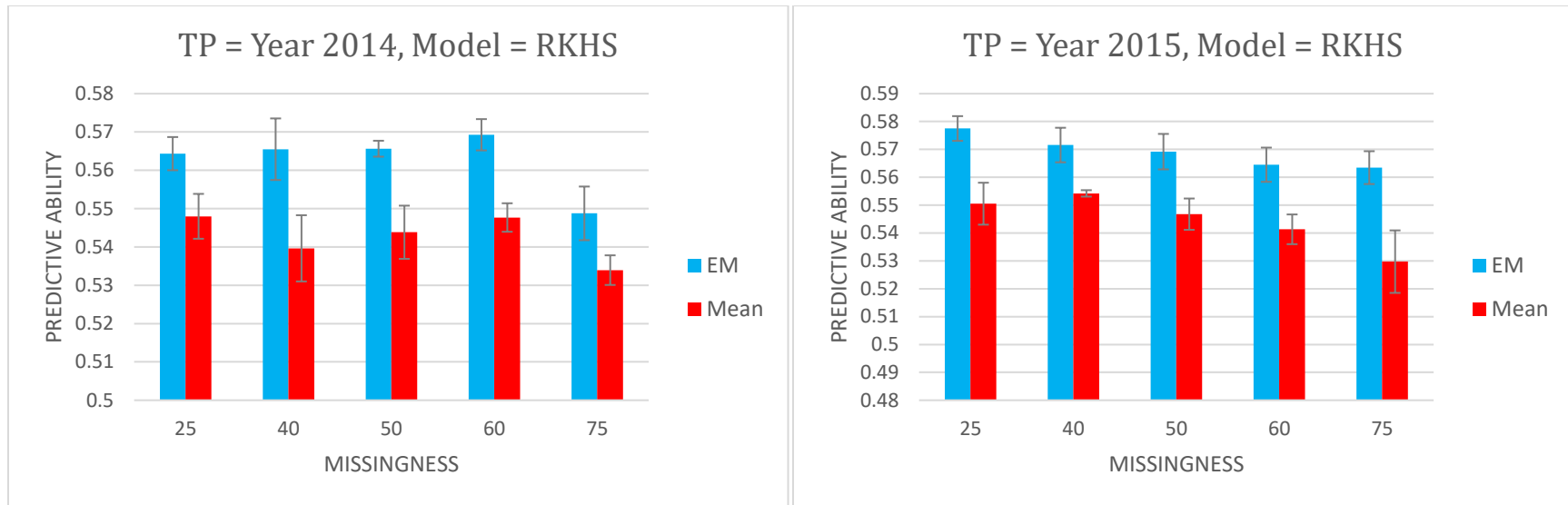


Figure 2 Comparison of two missing data imputation methods, EM and Mean, based on the predictive ability from RKHS cross-validation model (with SNP effect only) across a gradient of missing marker data ratio (Missingness); TP = training population; bandwidth parameter was set to 0.1 for all models.

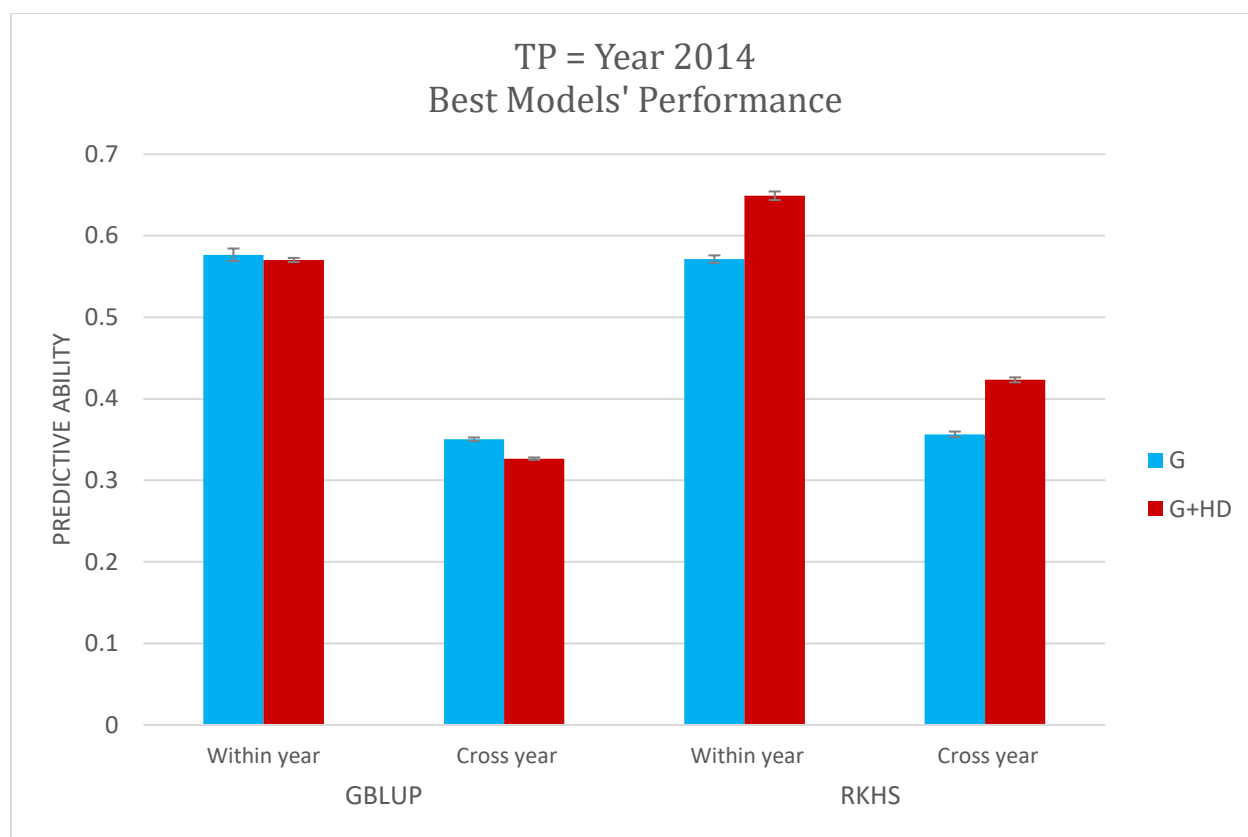


Figure 3 GBLUP and RKHS prediction accuracies from the best-performing models using individuals from year 2014 as training population (TP). Within each model type, within-year cross-validation predictive ability was compared to cross-year predictive ability. G = only marker effect and G+HD = marker effect and heading date as models' covariates.

Cross-year prediction using 2015 as training population

In general, cross year prediction results showed higher consistency when 2015 was used to predict 2014 (average predictive ability: 0.378 ± 0.014). Also, to predict grain yield in 2014, RKHS consistently performed better; even in cross-year prediction without covariates, RKHS outperformed the linear GBLUP model, as opposed to the slightly higher accuracy estimate obtained from GBLUP in the 2014 within-year cross-validation (0.577 and 0.571 for GBLUP and RKHS respectively). The best performance was obtained when RI_5 was included in the model as a covariate, though inclusion of both HD and rust ratings (RI_5) produced comparable results (Figure 4). The use of flowering time (HD) as a covariate in the prediction model is not recommended, as this model resulted in the lowest predictive ability (0.352) and the highest standard error (0.00663) of all cross-year predictions. Overall, when using 2015 as training population to predict 2014's yield data, decrease in the predictive ability of about 32 and 38% were observed for GBLUP and RKHS, respectively.

To summarize the predictive ability performance for our two consecutive years' yield data, results from RKHS, in general, produced higher accuracy than that of GBLUP (Figures 3 and 4). The RKHS also benefited considerably from the inclusion of covariates; GBLUP was at its best only when SNP markers were used, except for the very slight gain of predictive ability in the cross-year prediction when 2015's data was used to predict 2014's grain yield and HD was included as a covariate (0.374 versus 0.375). As a result, including covariates like HD and rust infection ratings was not recommended for GBLUP; in the case of within-year cross-validation using 2015 yield data, prediction performance for GBLUP was in fact at its lowest when covariates were included in the model (Figure 4).

Also, the conventional cross-validation used to evaluate model performance might result in an over-estimation of predictive ability; predictive ability for within-year cross-validation ranged from 0.533 (GBLUP, TP = 2015, SNP marker data + HD + rust rating) to 0.695 (RKHS, TP=2015, SNP marker data + HD + rust rating). Taking year effect into consideration, predictive ability was dramatically reduced, an evident decrease of 37% was observed.

To investigate the factors affecting model performance, we also compared the numbers of SNP markers and the missing data ratio to determine the genetic information content required for predictive analysis. Cross-year prediction models required larger number of SNP markers for 8 out of the 12 scenarios to achieve comparable prediction results, suggesting the complex genetic architecture of grain yield trait (Table 3). Finally, search for the optimal bandwidth parameter h across models failed to identify a single bandwidth value. The pattern changed with the training population, model composition, and the number of markers employed (Appendix A). According to our results, h values could vary between 0.1 and 1 is recommended for the acquisition of the highest predictive ability.

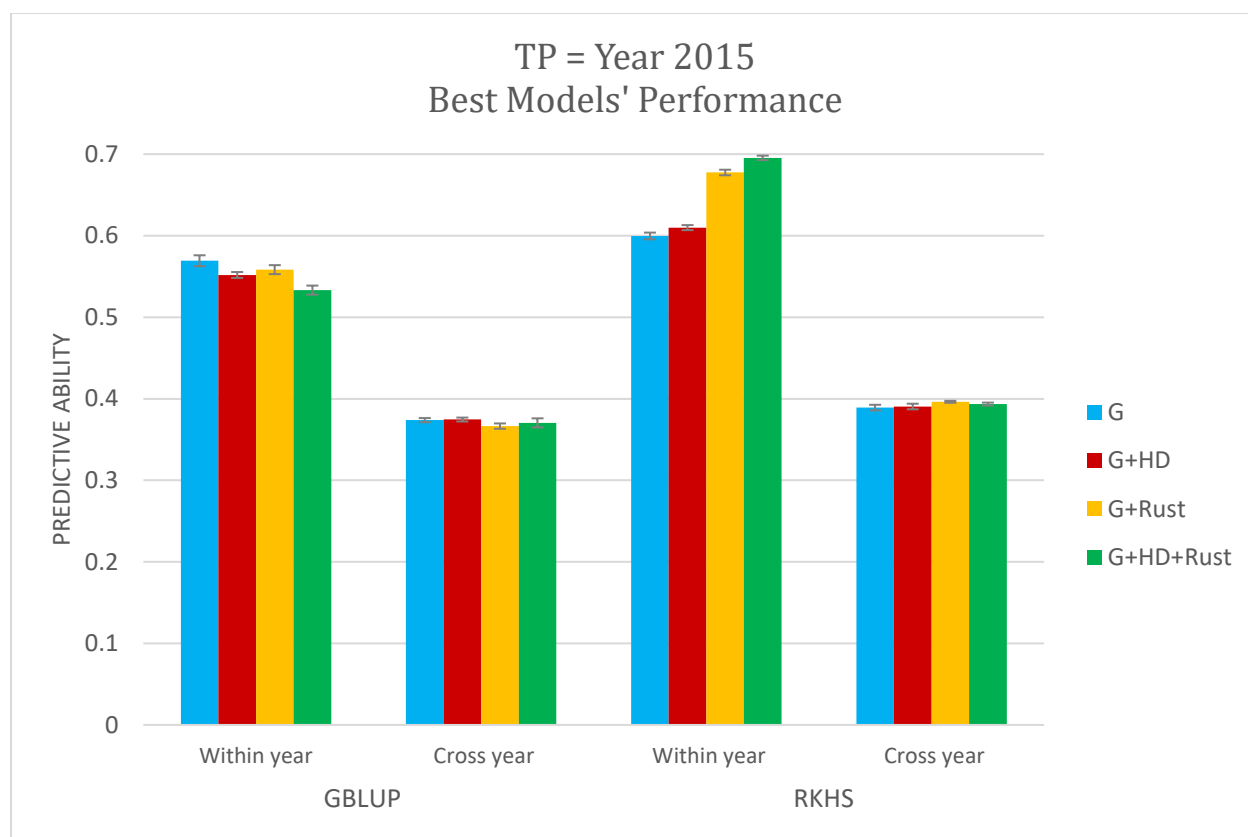


Figure 4 GBLUP and RKHS predictive ability from the best-performing models using individuals from year 2015 as training population (TP). Within each model type, within-year cross-validation predictive ability was compared to cross-year predictive ability. G = only marker effect; G+HD = marker effect and heading date as covariate; G+Rust = marker effect and disease index as covariate; G+HD+Rust = all three components in the model.

Marker selection

The best model for within-year cross-validation was the RKHS with TP = 2015, SNP marker data + HD + rust rating while the best model for cross-year prediction was also RKHS with TP = 2014, SNP marker data + HD and both models required the marker subset at call rate of 0.6, hence 5,726 SNPs were considered to be a reasonable starting point of our investigation on marker selection. Based on the whole population SNP marker data, number of SNPs remained after filtration by the absolute pairwise correlation value t is shown in Table 4. Since the filtration step was carried out within each fold, the pairwise correlation value was estimated based on the training set rather than the whole population. The numbers in Table 4 could be seen as a reference, while the actual number of SNPs varied with the changing training population.

Overall, a comparable pattern of predictive ability was observed for both within-year and cross-year models (Figure 5), where predictive ability remained constant until $t = 0.4$ and at $t = 0.3$, the prediction abilities from both within-year and cross-year models were reduced and showing a significant loss of information due to the sparse marker density. For the DH population, approximately 1,500 SNPs at the absolute pairwise correlation threshold of $t = 0.4$ could result in a similar predictive ability when the full set of 5,726 SNPs was used.

Table 3 Best-performing models and the number of SNPs required (SE = standard error).

Training population	Algorithm	Model	Predictive ability (\pm SE)		Number of SNPs (call rate)		
			<i>Within</i>	<i>Cross</i>	<i>Within</i>	<i>Cross</i>	
2014	<i>GBLUP</i>	<i>G</i>	0.577 (± 0.0077)	0.350 (± 0.0022)	5,726 (0.6)	7,260 (0.5)	
		<i>G+HD</i>	0.570 (± 0.0026)	0.327 (± 0.0015)	7,260 (0.5)	5,726 (0.6)	
	<i>RKHS</i>	<i>G</i>	0.571 (± 0.0045)	0.356 (± 0.0034)	5,726 (0.6)	7,260 (0.5)	
		<i>G+HD</i>	0.649 (± 0.0053)	0.423 (± 0.0030)	7,260 (0.5)	5,726 (0.6)	
	2015	<i>GBLUP</i>	<i>G</i>	0.569 (± 0.0067)	0.374 (± 0.0025)	4,010 (0.75)	9,244 (0.4)
			<i>G+HD</i>	0.552 (± 0.0036)	0.375 (± 0.0024)	5,726 (0.6)	7,260 (0.5)
<i>G+Rust</i>			0.558 (± 0.0055)	0.366 (± 0.0033)	5,726 (0.6)	4,010 (0.75)	
<i>G+HD+Rust</i>			0.533 (± 0.0056)	0.370 (± 0.0056)	4,010 (0.75)	7,260 (0.5)	
<i>RKHS</i>		<i>G</i>	0.600 (± 0.0041)	0.389 (± 0.0034)	4,010 (0.75)	7,260 (0.5)	
		<i>G+HD</i>	0.610 (± 0.0031)	0.391 (± 0.0034)	5,726 (0.6)	5,726 (0.6)	
	<i>G+Rust</i>	0.677 (± 0.0034)	0.396 (± 0.0012)	5,726 (0.6)	7,260 (0.5)		
	<i>G+HD+Rust</i>	0.695 (± 0.0028)	0.394 (± 0.0018)	5,726 (0.6)	7,260 (0.5)		

Table 4 Number of SNP markers within each correlation group based on whole population data

	Absolute pairwise correlation threshold t									
	Full Set	1.0	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2
# of SNPs	5,726	4,976	4,506	4,241	3,883	3,338	2,595	1,473	267	27

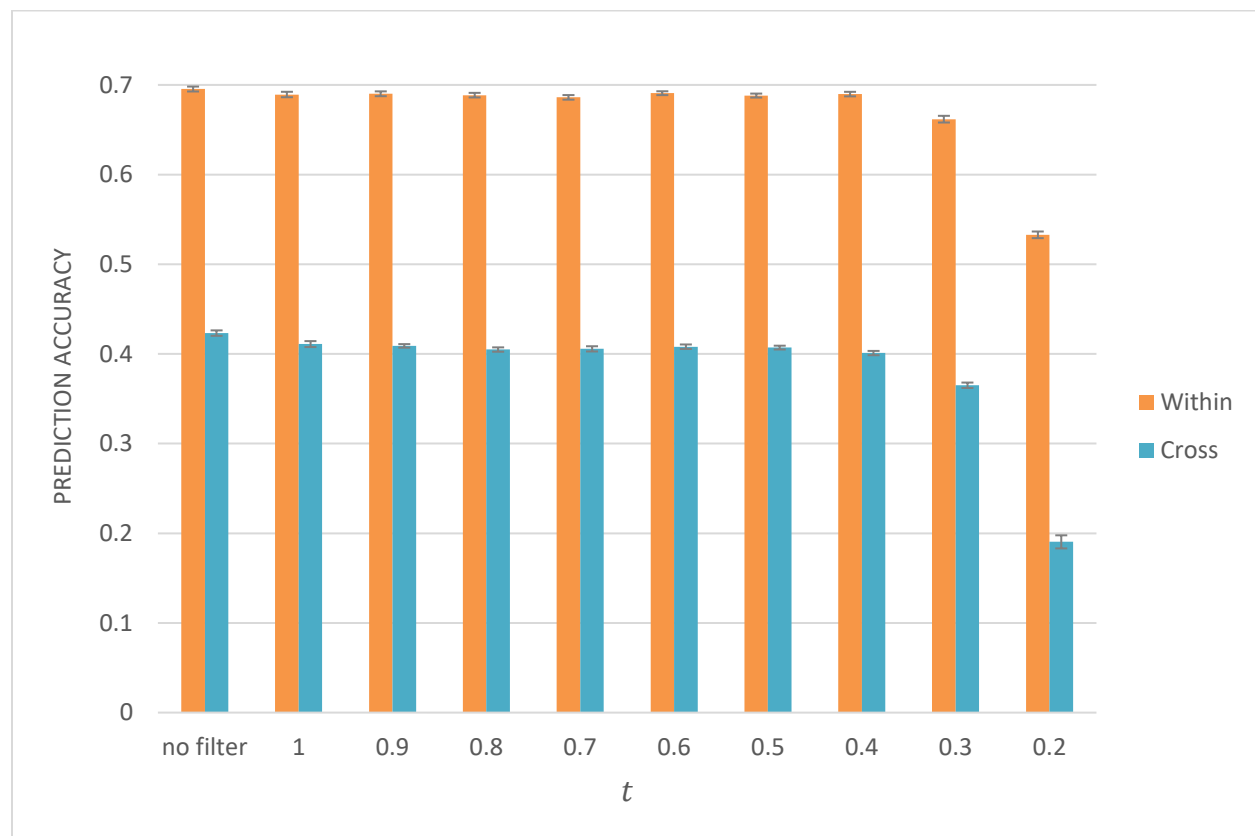


Figure 5 Predictive ability from the best within-year cross-validation model (Within: year 2015 RKHS model with marker effect and both heading date and disease index as covariates), and the best cross year prediction model (Cross: year 2014 predicting 2015 RKHS model with marker effect and heading date as covariate) across subsets of marker filtered by absolute pairwise correlation threshold t .

Consistency of elite line selection

The trend of average ranking distance (\bar{d}) has shown some substantial fluctuation among the top 30 lines, but the top 10% individuals produced consistent distance values under or close to 3, indicating that the selection of the best individuals was consistent across different environments (Figure 6). The number then gradually increased and plateaued at an average distance close to 6 when about 80% of the population were included, suggesting that evaluating the majority of the population with moderate performances was less certain relative to the top individuals. There was also a slight decrease in the average distance with inclusion of the lowest-ranking individuals, showing a steady assessment for those poorly-performing lines. In summary, the models were more consistent in selecting individuals with extreme performances than evaluating average lines.

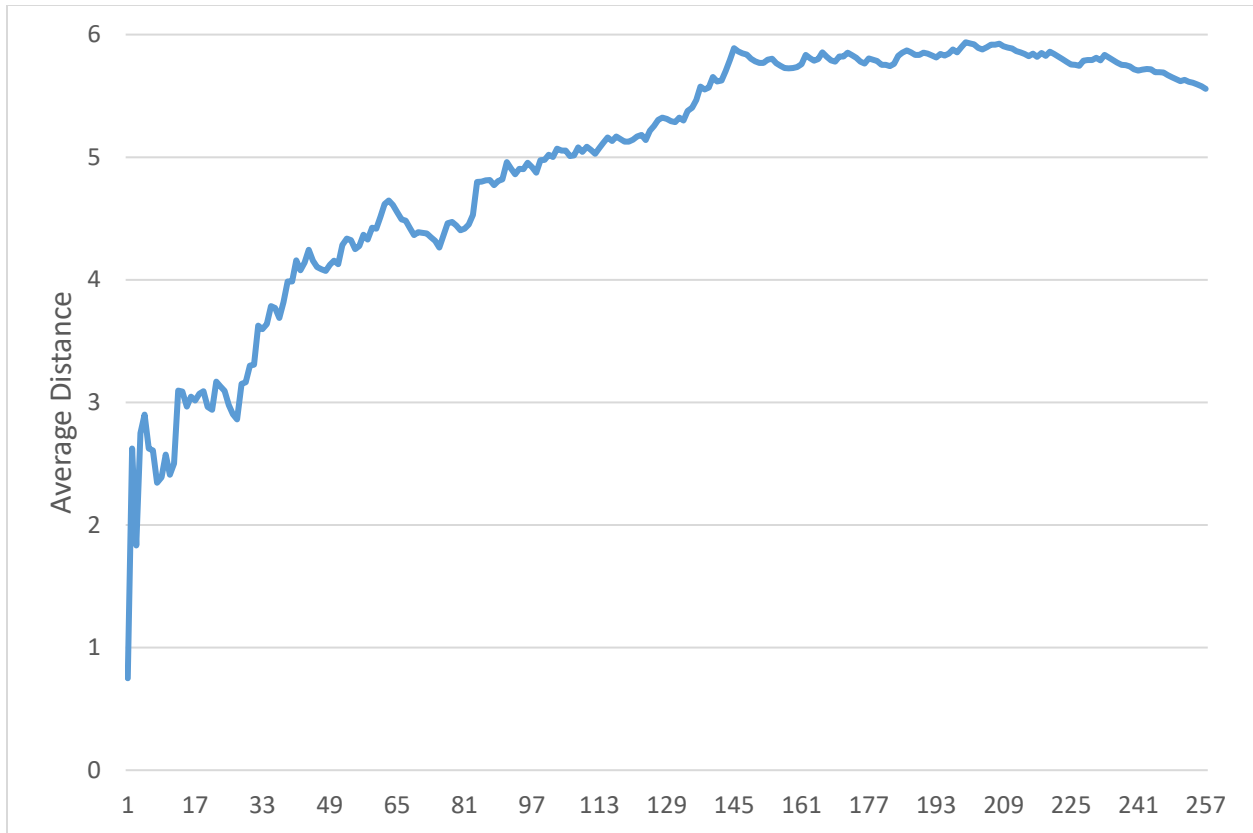


Figure 6 Trend of average ranking distance over the number of individuals considered. Analysis started with the ranking distance of the best-performing individual in the year 2014, proceeded by adding the next best individual's ranking distance and taking the mean, until all 257 individuals in the population were included

Discussion

Efficacy of genomic selection has been widely evaluated since its inception in 2001 (Meuwissen et al., 2001). Evidences for the method's potential in wheat breeding programs were demonstrated by a number of studies (e.g. Crossa et al., 2010; 2011; Poland et al., 2012; He et al., 2016; Huang et al., 2016; Michel et al., 2016; Saint Pierre et al., 2016). Among these studies, only few focused on the inter-year performance of models or considered the application in actual breeding programs. In this study, we assessed the predictive ability of genomic selection models using grain yield data from two successive years of a hard red winter wheat DH population as an example. Cross-year prediction using RKHS algorithm was found to be the most reliable method among all tested alternatives; we also demonstrated that consistent selection of lines with extreme values is achievable for the practicality.

Model comparison

In our study, the RKHS method showed equal or higher predictive ability than GBLUP, the linear additive alternative, for grain yield prediction of a winter wheat DH population. This observation corresponds to a number of previous studies that investigated genomic prediction model performance using grain yield components as targets. For example, Huang et al. (2016) reported similar accuracy differences of 0.33 between RKHS and GBLUP in predicting grain yield for 273 elite soft winter wheat lines. With a larger collection of 2,325 European elite winter wheat lines, He et al. (2016) attained 5% higher predictive ability that was associated with 17% reduction in standard error for RKHS than GBLUP when evaluating grain yield in multiple sites. Additionally, RKHS outperformed GBLUP including other methods such as BayesC π and artificial neural networks by 4% when predicting wheat grain yield (Heslot et al., 2012). Similar results were

reported by Crossa et al. (2010, 2011) where RKHS outperformed Bayesian LASSO (a similar algorithm to GBLUP but with marker-specific shrinkage) for grain yield prediction using 599 wheat lines and 94 elite spring wheat lines. Our results are in line with previously studies, highlighting the advantage of RKHS and suggesting its broader application in predicting polygenic, complex traits like grain yield.

The bandwidth parameter (h) in RKHS is used to control the rate of decay of the co-variance between genotypes. For cross-year predictions of a single DH population, in theory a single value of h could be expected, when there are no new recombination events between genotypes. Our results, however, found an inconclusive result for the bandwidth parameter (Appendix A). The unsuccessful search for a single optimal bandwidth parameter was also discussed in the original work that proposed the use of RKHS for genomic selection, as de los Campos et al. (2010) indicated that variation of the optimal value of h is expected when distribution of observed genetic distances changes, which in part could be due to the different numbers of SNP markers used in our study. Other factors such as the genetic architecture of trait of interest and choice of kernel function also affect estimate of this parameter (de los Campos et al., 2010). Cross-validation is commonly used as independent evaluation to identify the optimal value for the bandwidth (Härdle & Linton, 1994), alternatives like the Kernel Averaging method proposed by de los Campos et al. (2010) and Bayesian based selection of h in Pérez-Elizalde et al. (2015) can also be considered without going through a large number of grid search.

Although the superiority of RKHS diminished significantly (on average a 0.25 decrease in predictive ability) from within-year cross-validation to cross-year prediction, in our study, this decrease in accuracy was present in both parametric and non-parametric algorithms for all scenarios (Figures 3 and 4). Such inflation of predictive ability, plausibly disadvantageous for

evaluating GS applicability using cross-validation, as this could be the result of the common environmental variation (Lorenz et al., 2011). In most of the literatures that estimate predictive ability based on cross-validations, Pearson product-moment correlation between GEBV and true breeding value (TBV), $r(\text{GEBV}, \text{TBV})$, was used to reflect the confidence of how GEBV can be used to replace field evaluation. Since TBV is unknown, and we could only measure the observed phenotype (Obs), and evaluation of model performance is based on $r(\text{GEBV}, \text{Obs})$, which is assumed to be the product between $r(\text{GEBV}, \text{TBV})$ and $r(\text{Obs}, \text{TBV})$. This assumption is only valid when the common element between GEBV and Obs is just TBV, and, more importantly, the assumption of uncorrelated error terms between GEBV and Obs also needs to remain true. It can be expected that having both training and validation sets evaluated in the same environment in the same year constitutes a violation to the assumption of uncorrelated error terms, additionally the presence of $G \times E$ is expected to produce an upward bias in predictive ability for within year cross-validation.

On average, we found 32 and 38% decrease in accuracy when switching from within year cross-validation to cross-year prediction using GBLUP and RKHS, respectively. Michel et al. (2016) also observed major decline in predictive ability for cross-year prediction in comparison to within-year cross-validation in a 5-year study for 659 commercial winter wheat lines. Similarly, an average accuracy drop from 0.65 to 0.5 was found by He et al. (2016) when switching from within year cross-validation to cross-year prediction for European elite winter wheat tested in two successive years; also, a larger decrease of 50% in predictive ability was reported in a two-generation sugar beet study (Hofheinz et al., 2012). When evaluating applicability of GS, all these studies, including the present research, reckon that cross-year prediction should be considered to reduce the upwards bias in prediction models.

Further, the RKHS method had on average higher degree of overfitting in within-year cross-validation comparing to GBLUP (RKHS 0.253 vs. GBLUP 0.182). The higher degree of variability of RKHS was also observed by Heslot et al. (2012) and thought to likely be due to model overfitting. The strength of RKHS in capturing genetic effects, including the high-level interaction terms has been recognized by Gianola & Kaam (2008). In k-fold cross-validation, the evaluation of model performance can be broken down into bias and variance components. While unbiasedness is cited as the beneficial quality of a model, low variance is just as important. The observed error in RKHS, at least in our winter wheat DH population, is suggestive that the kernel used to capture genomic relationship amongst individuals did not fully encompass Mendelian sampling term in between training and validation populations. Though overall predictability of GBLUP was slightly lower, the reduced errors in GBLUP model might be indicative of a common genetic architecture shared between training and validation in k-fold cross-validation preserved by the extensive linkage disequilibrium (LD) blocks in DH populations.

The rust infection in 2015 might have caused a strong G x E effect to be picked up by cross-validation, as the model (TP=2015, RKHS, G+HD+Rust) which produced the highest within-year predictive ability did not perform as well in the cross-year prediction, and the highest cross-year predictive ability was actually found using year 2014 as training population. This result is in accordance with Saint Pierre et al. (2016)'s finding that the highest predictive ability was from an environment without the presence of any dominant biotic or abiotic stresses. Hence we would advocate using individuals evaluated in a stable year in regard to environmental status as training population for genomic selection.

Missing data imputation

The comparison between mean and EM imputation methods indicated that the latter consistently produced better model performance. Although our finding regarding EM method's superiority agrees with Poland et al. (2012), their study found only lower imputation error from EM than population mean for the masked non-missing genotypes, with no advantage of EM imputation in predictive ability for yield was observed relative to mean imputation. Possible reasons for this discrepancy could be: 1) the relatedness between the training and validation population (Poland et al. (2012) study had no common full-sib lines between the training and validation set, while every individual in our analyses shared the same pair of parents); 2) the LD between the correctly imputed markers and the QTL; in other words, if the missing values of SNPs in high LD with QTL were imputed with precision, imputation would lead to improvement of the predictive ability of the model, which could likely be our case. Otherwise, this improvement might not be obvious when the majority of the imputed markers is distant from the underlying QTLs of the target traits. In summary, our results confirmed the superiority of EM over mean imputation for DH populations where progeny are in moderate to high correlation.

Marker selection

The unprecedented efficiency of next-generation sequencing technology has created a paradigm-shift that changes genetic research from trait-driven science to genetic-driven discovery. Accompanied with this rapid advancement, issues in data-information inequality has become increasingly important as “information volume” is often smaller than “data volume”. A simulation study of dairy cattle and corn breeding showed that accuracies of prediction first increased with number of SNPs, then plateaued in spite of the growing quantity of markers (Habier et al., 2013).

In another study of a closely-related wheat population the authors postulated a comparable performance of 1,827 GBS SNP markers relative to 34,749 SNPs (Poland et al., 2012); similar predictability for wheat grain yield in Crossa et al. (2010) and de los Campos et al. (2009) was later achieved in 2011 with fewer genetic markers, showed in Crossa et al. (2011). Using a cross-environment validation, our results support, to approach comparable level of predictive ability in grain yield of a winter wheat DH population, that only a moderate number of SNP markers are required. Such lack of improvement with additional data points is not only the resource for inefficiency, but also the underlying cause of correlated errors. The level of LD in the population and the relatedness among individuals are the two main contributors to when the plateau will take place (de los Campos et al., 2013). With our DH population, the long spans of LD in the genome and high relatedness within the training population and between training and validation populations, filtering the SNP markers based on their correlation coefficients could produce satisfactory predictive ability while requiring less computational effort and time.

Conclusion

Over the last decade, a substantial amount of effort has been dedicated to exploring and evaluating the applicability of genomic selection for crop improvement. With a more realistic two-generation validation, our result verified the superiority of RKHS method over the whole-genome regression GBLUP; the observed slightly larger errors in RKHS is mainly due to model overfitting suggesting the presence of considerable Mendelian sampling terms even within a DH population. Model performance evaluation based on within-year cross-validation is likely to be biased, and when aiming to shorten breeding cycles in the line development stage of a wheat breeding program, a more ideal design like our two-generation validation should be considered with multi-location field data to handle correlated errors.

Up to this date, only a few investigated the realized outcome from GS in the context of an ongoing breeding program. Among these, our study demonstrated that the confidence of line selection based on genomic selection could be achieved. Selection of lines encompassing high breeding values with precision is considered to be a prerequisite to selective breeding endeavor (Blondel et al., 2015). Given the considerable differences in the predictive abilities from the various models examined in the present, forward selection for high performer was consistent and the ranking differential was small, even with a moderate number of SNP markers. Though the differential was slightly larger, rankings of the low performers were also considered stable. In summary, the robust assessment in line selection advocates the advantage of implementing genomic selection in wheat variety development.

Bibliography

- Alexandratos, N. (2009). World food and agriculture to 2030/50. *Highlights and Views from Mid*.
Retrieved from <http://ss.rrojasdatabank.info/ak969e00.pdf>
- Baenziger, P. S., & Depauw, R. M. (2009). Wheat Breeding: Procedures and Strategies. In B. F. Carver (Ed.), *Wheat Science and Trade* (pp. 273–308). Oxford, UK: Wiley-Blackwell.
Retrieved from <http://doi.wiley.com/10.1002/9780813818832.ch13>
- Bassi, F. M., Bentley, A. R., Charmet, G., Ortiz, R., & Crossa, J. (2016). Breeding schemes for the implementation of genomic selection in wheat (*Triticum* spp.). *Plant Science*, 242, 23–36.
<https://doi.org/10.1016/j.plantsci.2015.08.021>
- Bernardo, R., & Yu, J. (2007). Prospects for Genomewide Selection for Quantitative Traits in Maize. *Crop Science*, 47(3), 1082. <https://doi.org/10.2135/cropsci2006.11.0690>
- Blondel, M., Onogi, A., Iwata, H., & Ueda, N. (2015). A Ranking Approach to Genomic Selection. *PLOS ONE*, 10(6), e0128570. <https://doi.org/10.1371/journal.pone.0128570>
- Crossa, J., Campos, G. de los, Pérez, P., Gianola, D., Burgueño, J., Araus, J. L., ... Braun, H.-J. (2010). Prediction of Genetic Values of Quantitative Traits in Plant Breeding Using Pedigree and Molecular Markers. *Genetics*, 186(2), 713–724.
<https://doi.org/10.1534/genetics.110.118521>
- Crossa, J., Pérez, P., Campos, G. de los, Mahuku, G., Dreisigacker, S., & Magorokosho, C. (2011). Genomic Selection and Prediction in Plant Breeding. *Journal of Crop Improvement*, 25(3), 239–261. <https://doi.org/10.1080/15427528.2011.558767>
- de los Campos, G., Gianola, D., Rosa, G. J. M., Weigel, K. A., & Crossa, J. (2010). Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert

- spaces methods. *Genetics Research*, 92(4), 295–308.
<https://doi.org/10.1017/S0016672310000285>
- de los Campos, G., Hickey, J. M., Pong-Wong, R., Daetwyler, H. D., & Calus, M. P. L. (2013). Whole-Genome Regression and Prediction Methods Applied to Plant and Animal Breeding. *Genetics*, 193(2), 327–345. <https://doi.org/10.1534/genetics.112.143313>
- de los Campos, G., Naya, H., Gianola, D., Crossa, J., Legarra, A., Manfredi, E., ... Cotes, J. M. (2009). Predicting Quantitative Traits with Regression Models for Dense Molecular Markers and Pedigree. *Genetics*, 182(1), 375–385.
<https://doi.org/10.1534/genetics.109.101501>
- Edwards, J. T., Hunger, R. M., Smith, E. L., Horn, G. W., Chen, M.-S., Yan, L., ... Carver, B. F. (2012). “Duster” Wheat: A Durable, Dual-Purpose Cultivar Adapted to the Southern Great Plains of the USA. *Journal of Plant Registrations*, 6(1), 37.
<https://doi.org/10.3198/jpr2011.04.0195crc>
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLOS ONE*, 6(5), e19379.
<https://doi.org/10.1371/journal.pone.0019379>
- Endelman, J. B. (2011). Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *The Plant Genome Journal*, 4(3), 250.
<https://doi.org/10.3835/plantgenome2011.08.0024>
- Gianola, D., & Kaam, J. B. C. H. M. van. (2008). Reproducing Kernel Hilbert Spaces Regression Methods for Genomic Assisted Prediction of Quantitative Traits. *Genetics*, 178(4), 2289–2303. <https://doi.org/10.1534/genetics.107.084285>

- Härdle, W., & Linton, O. (1994). Chapter 38 Applied nonparametric methods. In B.-H. of Econometrics (Ed.) (Vol. 4, pp. 2295–2339). Elsevier. Retrieved from <http://www.sciencedirect.com/science/article/pii/S1573441205800078>
- He, S., Schulthess, A. W., Mirdita, V., Zhao, Y., Korzun, V., Bothe, R., ... Jiang, Y. (2016). Genomic selection in a commercial winter wheat population. *Theoretical and Applied Genetics*, 129(3), 641–651. <https://doi.org/10.1007/s00122-015-2655-1>
- Heslot, N., Yang, H.-P., Sorrells, M. E., & Jannink, J.-L. (2012). Genomic Selection in Plant Breeding: A Comparison of Models. *Crop Science*, 52(1), 146. <https://doi.org/10.2135/cropsci2011.06.0297>
- Heslot, N., Jannink, J.-L., & Sorrells, M. E. (2015). Perspectives for Genomic Selection Applications and Research in Plants. *Crop Science*, 55(1), 1–12. <https://doi.org/10.2135/cropsci2014.03.0249>
- Hofheinz, N., Borchardt, D., Weissleder, K., & Frisch, M. (2012). Genome-based prediction of test cross performance in two subsequent breeding cycles. *Theoretical and Applied Genetics*, 125(8), 1639–1645. <https://doi.org/10.1007/s00122-012-1940-5>
- Huang, M., Cabrera, A., Hoffstetter, A., Griffey, C., Sanford, D., Costa, J., ... Sneller, C. (2016). Genomic selection for wheat traits and trait stability. *Theoretical and Applied Genetics*, 1–14. <https://doi.org/10.1007/s00122-016-2733-z>
- Hunger, R. M., Edwards, J. T., Bowden, R. L., Yan, L., Rayas-Duarte, P., Bai, G., ... Carver, B. F. (2014). “Billings” Wheat Combines Early Maturity, Disease Resistance, and Desirable Grain Quality for the Southern Great Plains, USA. *Journal of Plant Registrations*, 8(1), 22. <https://doi.org/10.3198/jpr2012.11.0053crc>

- Kiseleva, A. A., Shcherban, A. B., Leonova, I. N., Frenkel, Z., & Salina, E. A. (2016). Identification of new heading date determinants in wheat 5B chromosome. *BMC Plant Biology*, 16(1), 35–46. <https://doi.org/10.1186/s12870-015-0688-x>
- Li, G., Wang, Y., Chen, M.-S., Edae, E., Poland, J., Akhunov, E., ... Yan, L. (2015). Precisely mapping a major gene conferring resistance to Hessian fly in bread wheat using genotyping-by-sequencing. *BMC Genomics*, 16, 108. <https://doi.org/10.1186/s12864-015-1297-7>
- Lorenz, A. J., Chao, S., Asoro, F. G., Heffner, E. L., Hayashi, T., & Iwata, H. (2011). 2 Genomic Selection in Plant Breeding: Knowledge and Prospects. *Advances in Agronomy*, 110, 77.
- Meuwissen, T. H. E., Hayes, B. J., & Goddard, M. E. (2001). Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics*, 157(4), 1819–1829.
- Michel, S., Ametz, C., Gungor, H., Epure, D., Grausgruber, H., Löschenberger, F., & Buerstmayr, H. (2016). Genomic selection across multiple breeding cycles in applied bread wheat breeding. *Theoretical and Applied Genetics*, 1–11. <https://doi.org/10.1007/s00122-016-2694-2>
- Opgen-Rhein, R., & Strimmer, K. (2007). Accurate ranking of differentially expressed genes by a distribution-free shrinkage approach. *Statistical Applications in Genetics and Molecular Biology*, 6(1). Retrieved from [http://www.degruyter.com/dg/viewarticle/j\\$002fsagmb.2007.6.1\\$002fsagmb.2007.6.1.1252\\$002fsagmb.2007.6.1.1252.xml](http://www.degruyter.com/dg/viewarticle/j$002fsagmb.2007.6.1$002fsagmb.2007.6.1.1252$002fsagmb.2007.6.1.1252.xml)
- Pérez-Elizalde, S., Cuevas, J., Pérez-Rodríguez, P., & Crossa, J. (2015). Selection of the Bandwidth Parameter in a Bayesian Kernel Regression Model for Genomic-Enabled

- Prediction. *Journal of Agricultural, Biological, and Environmental Statistics*, 20(4), 512–532. <https://doi.org/10.1007/s13253-015-0229-y>
- Poland, J. A., Brown, P. J., Sorrells, M. E., & Jannink, J.-L. (2012). Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-by-Sequencing Approach. *PLOS ONE*, 7(2), e32253. <https://doi.org/10.1371/journal.pone.0032253>
- Poland, J., Endelman, J., Dawson, J., Rutkoski, J., Wu, S., Manes, Y., ... Jannink, J.-L. (2012). Genomic Selection in Wheat Breeding using Genotyping-by-Sequencing. *The Plant Genome Journal*, 5(3), 103. <https://doi.org/10.3835/plantgenome2012.06.0006>
- Saint Pierre, C., Burgueño, J., Crossa, J., Fuentes Dávila, G., Figueroa López, P., Solís Moya, E., ... Singh, S. (2016). Genomic prediction models for grain yield of spring bread wheat in diverse agro-ecological zones. *Scientific Reports*, 6. <https://doi.org/10.1038/srep27312>
- Schäfer, J., & Strimmer, K. (2005). A Shrinkage Approach to Large-Scale Covariance Matrix Estimation and Implications for Functional Genomics. *Statistical Applications in Genetics and Molecular Biology*, 4(1), 1–32.
- Tadesse, W., Rajaram, S., Ogbonnaya, F. ., Sanchez-Garcia, M., Sohail, Q., & Baum, M. (2016). Wheat. In M. Singh & S. Kumar (Eds.), *Broadening the Genetic Base of Grain Cereals* (pp. 9–26). New Delhi: Springer India. Retrieved from <http://link.springer.com/10.1007/978-81-322-3613-9>
- United States Department of Agriculture. (2016a). *Grain: World Markets and Trade*. Retrieved from <https://www.fas.usda.gov/>
- United States Department of Agriculture. (2016b). *World Agricultural Supply and Demand Estimates*. Retrieved from <http://www.usda.gov/oce/commodity/wasde/>

VanRaden, P. M. (2008). Efficient Methods to Compute Genomic Predictions. *Journal of Dairy Science*, 91(11), 4414–4423. <https://doi.org/10.3168/jds.2007-0980>

Appendices

Appendix A Effect of bandwidth parameter h on model predictive ability across training population, model composition, and number of markers employed

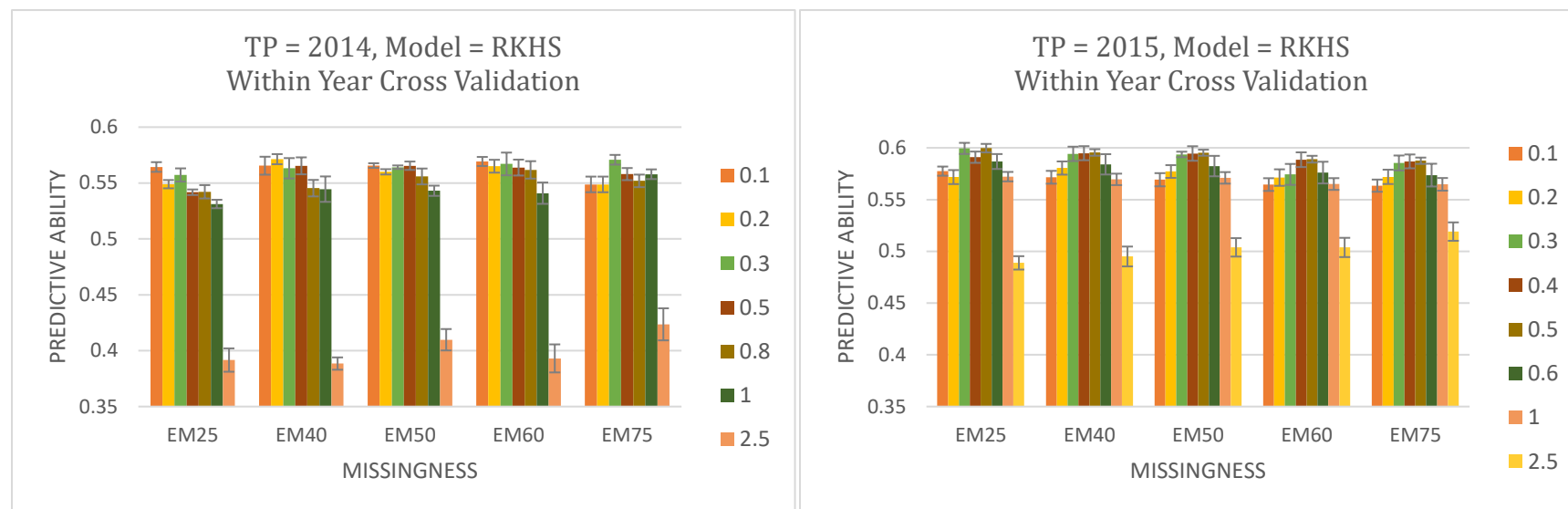


Figure 7 Comparison of bandwidth parameter h based on predictive ability from RKHS (with year 2014 and 2015 as training population respectively) cross validation model (with SNP effect only) across marker missingness levels

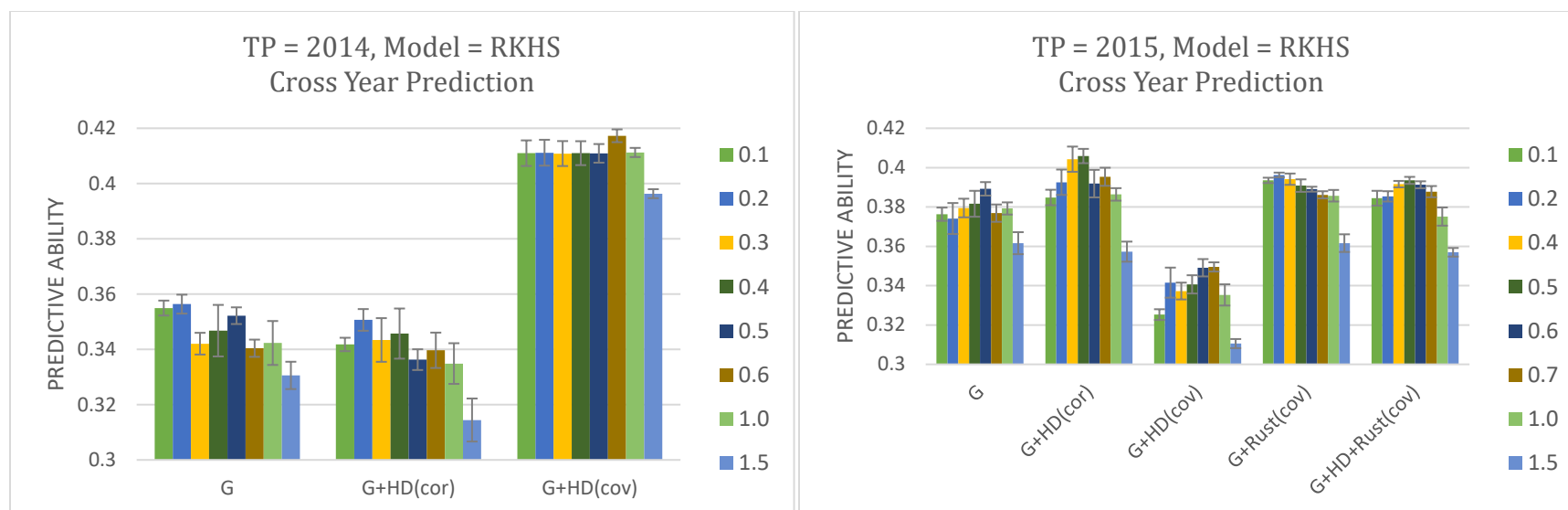


Figure 8 Comparison of bandwidth parameter h based on predictive ability from RKHS cross year prediction model (with year 2014 and 2015 as training population respectively) across different model components: G=only marker effect in the model; G+HD(cor)=marker effect in the model with heading date-corrected phenotype as response variable; G+HD(cov)=marker effect and heading date as covariate in the model; G+Rust(cov)=marker effect and disease index as covariate in the model; G+HD+Rust(cov)=marker effect and both heading date and disease index as covariates in the model