




MMGS: a novel genomic prediction framework to integrate genotype, environment and their interactions for multi-environment breeding trials

Mingjia Zhu¹, Zeyu Zheng^{1,2}, Wei Liu³, Yu Han³ , Wenjie Mou⁴, Tongming Yin⁵, Xiaogang Dai⁵, Huaitong Wu⁵, Yongzhi Yang¹ , Yanjun Zan^{6,7,*}  and Jianquan Liu^{1,*}

¹State Key Laboratory of Herbage Innovation and Grassland Agro-Ecosystem, College of Ecology, Lanzhou University, Lanzhou, China

²Yazhouwan National Laboratory (YNL), Sanya, Hainan, China

³College of Life Sciences, China Sichuan University, Chengdu, Sichuan, China

⁴Departamento de Ciencias Agrarias y del Medio Natural, Escuela Politecnica Superior de Huesca, Universidad de Zaragoza, Huesca 22071, Spain

⁵College of Forestry, Nanjing Forestry University, Nanjing, Jiangsu, China

⁶Integrated Science Lab, Department of Plant Physiology, Umeå Plant Science Center, Umeå University, Umeå 90736, Sweden

⁷Tobacco Research Institute, Chinese Academy of Agricultural Sciences, Qingdao, China

*Corresponding authors. E-mail: liujq@nwipb.ac.cn; zanyanjun@caas.cn

Abstract

Accurately predicting the performance of trees and crops across diverse and changing climates is essential for matching genotypes to both current and future environments. Yet modelling the complex interplay among genotype, environment, and phenotype in multi-environment trials remains a major challenge. Here, we introduce a unified framework, polygenic environmental interaction (PEI), directly models genotype-by-environment interactions through integrating genotypes and environmental covariates. We implemented an ensemble of 15 estimators spanning parametric, non-parametric, and machine-learning approaches. We then benchmarked our framework against the classical reaction norm (RN) using three genetically distinct populations and three traits with variable genetic architectures. Furthermore, we released an open-source R package, Multiple-environments genomic selection (MMGS), on GitHub. Together, our study offers a flexible and computationally efficient approach for multi-environment genomic prediction, enhancing breeding efficiency, providing deeper insights into modelling the genotype-environment-phenotype continuum.

Introduction

Both artificially cultivated trees and herbaceous crops exhibited plastic responses to changes in their surrounding environments, with one genotype performing well in one environment yet underperforming in another [1–3]. This variability poses a significant challenge in breeding crops and trees to meet the growing global food demand amid climate change [4]. To address this, breeders routinely screen for elite plants by subjecting germplasm and breeding materials to multiple-environment trials (METs) [5, 6]. Despite the advancements achieved through METs, several challenges persist, including the extensive number of genotypes tested in preliminary trials, the high associated costs and limited seed availability [7].

Over the past decades, genomic prediction has been increasingly applied to assess the overall performance and stability of genotypes across diverse environments, yet accurately capturing their overall stability remains challenging [8–10]. Through METs, it has been observed that variation in traits related to fitness and productivity is influenced by complex interactions between

genotypes and the environment ($G \times E$) [11–15]. In the absence of genetic mechanisms underlying variation in complex traits across environments, reaction norm (RN) assumes each genotype responds to changes in environment with a unique intercept and slope, representing distinct biological mechanisms [16–18]. Building upon this framework, an additional random effect was added to capture interactions between high-dimensional markers and environmental covariates (ECs) [19]. This extension integrated genetic relatedness into the estimation of regression slope and intercept, avoiding assuming each genotype is independent [20, 21]. Recently, an environment index was included in the RN framework, and Ridge regression best linear unbiased prediction (rr-BLUP) was used to predict the regression slope and intercept. With these advancements, environment-specific predictions were achieved and prediction accuracy was significantly improved [20, 22]. However, further improvements are still needed to enhance the prediction of the cross-environment METs [23–25]. For example, modelling phenotypes in multiple environments through the RN framework assumes a linear relationship between genotypes and environments, which is not always supported

Received: 10 September 2025. Accepted: 27 January 2026. Published: 16 March 2026. Corrected and Typeset: 1 May 2026

© The Author(s) 2026. Published by Oxford University Press on behalf of the Nanjing Agricultural University.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

by empirical data [21, 26]. In addition, recent studies have revealed a polygenic basis of phenotype plasticity, underpinned by interactions from core genes, polygenic background, and their interactions with environments. These advancements in how interaction between genotypes and environment co-determine the variation of complex traits have not yet been implemented for predicting complex traits across the environment. Moreover, previous frameworks only employed rr-BLUP for predicting parametric estimators within the RN framework. The vast number of parametric or non-parametric estimators, such as Bayesian, non-linear, and machine learning methods, have not been fully integrated into cross-environment predictions to account for variation in the underlying genetic basis of targeted traits [19, 22, 27].

To address these challenges, we implemented a polygenic environmental interaction (PEI) framework, which generalizes the factorial regression models by directly modelling $G \times E$. Unlike existing frameworks, PEI explicitly integrates the Critical Environmental Regressor through Informed Search (CERIS) algorithm to identify the most informative ECs and incorporates a flexible ensemble of 15 parametric and non-parametric estimators (e.g. Bayesian, kernel, and machine learning estimators). This integration enables the framework to account for polygenic and potentially non-linear environmental responses while maintaining interpretability and computational tractability. To evaluate the performance of PEI, we compared our framework with the RN to evaluate the prediction accuracy and computational efficiency across three genetically and phenotypically diverse populations: an inbred wheat population [28], a Complete-diallel design plus Unbalanced Breeding-like Inter-Cross (CUBIC) maize population [29] and an F_1 half-sib willow population. Furthermore, we released an open-source R package (multiple-environments genomic selection [MMGS]), available at GitHub (<https://github.com/Ryougi-yukiro/MMGS>), to facilitate its application in breeding programs. Overall, the PEI framework and MMGS package provide a unified, flexible, and computationally efficient platform that bridges classical $G \times E$ modelling and modern machine learning approaches, enhancing our capacity to predict genotype performance and stability across diverse and future environments.

Results

Two types of frameworks for cross-environment prediction in METs

Previously, Jarquin *et al.* developed a framework for cross-environment genomic prediction by extending the RN framework with additional random effects to handle interactions between high-dimensional markers and ECs [18]. By accounting for genetic relatedness and environmental similarity, this framework significantly increased cross-environment genomic prediction accuracy. Building upon these progresses, a framework named CERIS-JRGA was developed [20, 22]. This framework includes four steps: (i) Applying the CERIS algorithm to identify an environmental index that captures the largest proportion of phenotypic variation. (ii) Regressing the observed phenotypes onto the identified environmental index to obtain an intercept and a slope estimate for each tested genotype. (iii) Treating intercept and slope as new 'traits' and performing genomic

prediction via ridge regression to predict the intercept and slope for each untested genotype. (iv) Predicting phenotypes of the untested genotypes using the predicted intercept, slope and the environmental index (Fig. 1a).

Here, we implemented a second framework directly modelling the effects from genetic markers, environmental factors, and their interactions. This framework is based on recent insights into the polygenic basis of phenotype plasticity and directly modelling interaction between polygenic genetic background and environmental factors (PEI) [22, 30, 31]. Consistent with the RN framework, the PEI starts by identifying the key environmental index that best captures the phenotypic variation (Fig. 1b). We subsequently constructed a relationship matrix that includes not only block diagonal matrices assembled from genetic markers but also all pairwise interactions between these markers and the identified environmental index.

The RN framework employed a mixed linear model to transform multi-environmental phenotypic records into RN parameters, while the PEI framework directly models these measures by integrating genotype, environment index, and their interactions (Fig. 1). Consequently, the PEI framework fully leverages the interactions between genotypes and environmental factors, offering a more comprehensive approach to capturing genotype-environment dynamics.

Parametric estimators for both frameworks could be performed in the linear mixed model or using various machine learning approaches. We therefore implemented 15 estimators, with unique assumptions on underlying genetic architecture and variable computational efficacy. Overall, these estimators could be classified into three major categories: parametric, semi-parametric, and non-parametric [32]. The parametric estimators include mixed linear models, such as genomic best linear unbiased prediction (G-BLUP) [33], BayesA (BA), BayesB (BB) [34], BayesC (BC) [35], Bayesian ridge regression (BRR) [36], Bayesian LASSO (BL) [37], least absolute shrinkage and selection operator (LASSO) [38], ridge regression (RR) [39], rr-BLUP [34], and elastic net (EN) [40]. The semi-parametric estimators include the reproducing kernel Hilbert space (RKHS) and multiple kernel RKHS (MKRKHS) [41]. The non-parametric estimators include support vector machine (SVM) [42], random forest (RF) [43], and gradient boosting machine (GBM) [44, 45]. These frameworks and estimators were implemented in an R package named as MMGS, which has been publicly available at GitHub, along with a detailed tutorial summarized in Fig. S1.

Hereafter, we refer to the two prediction frameworks, RN and PEI, as frameworks, and various parameter estimators as estimators. Since prediction performance may vary depending on the population structure of interested population (i.e. inbred/outbred cross, generation, and number of founders), the genetic basis of targeted trait, the number of environments, parameter estimators used, etc., we selected multiple traits from a maize multi-parental advanced inter-cross population, a wheat inbred population and a willow F_1 hybrid population to benchmark the prediction performance. For each population, three traits with varying genetic architecture were analysed with all 15 estimators. Prediction accuracy at each environment was calculated as average accuracy from 10-fold cross-validation with five replications. The averaged accuracy of all tested environments was used as an indicator of overall performance.

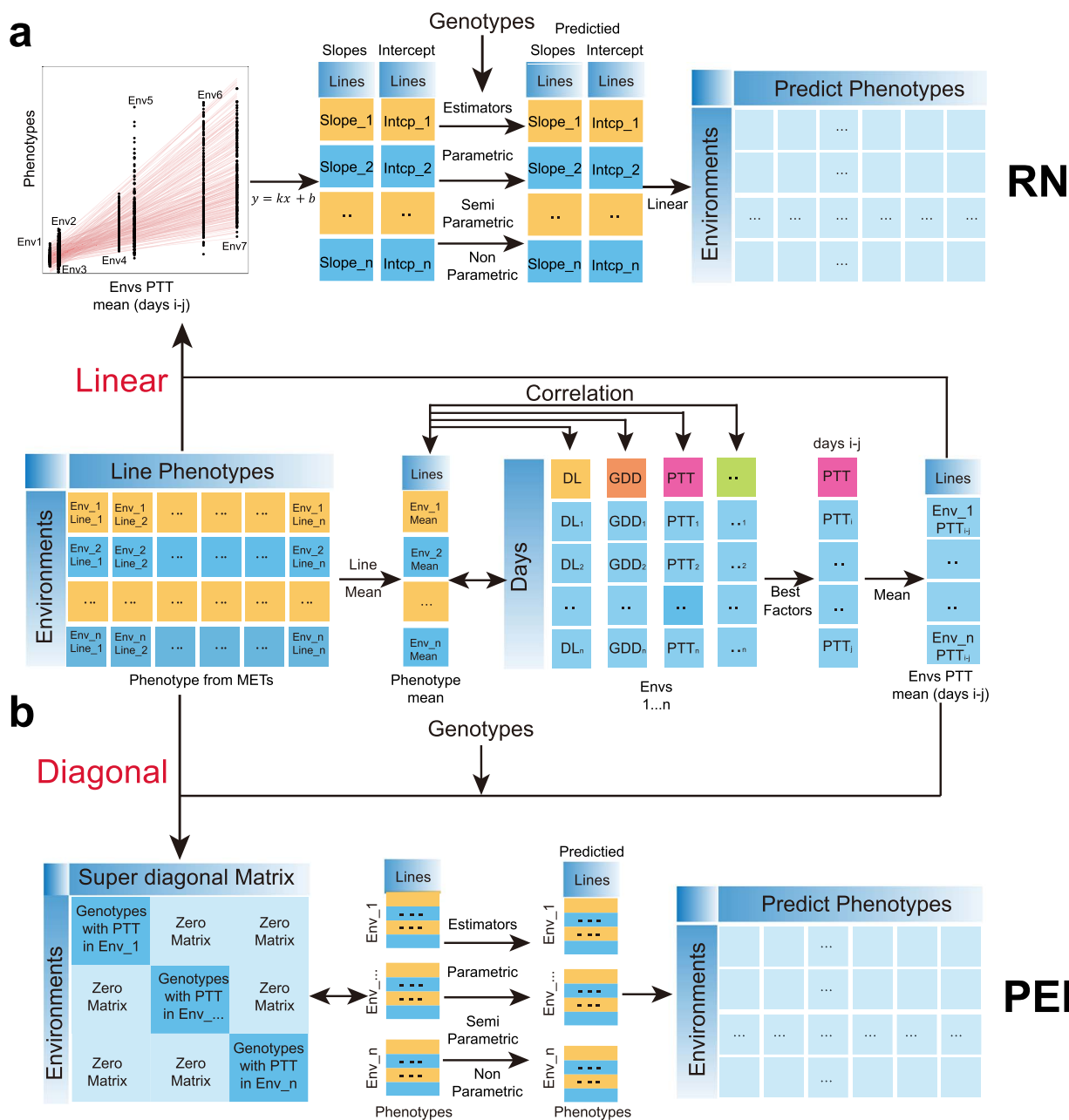


Figure 1 A schematic illustration of the two types of frameworks. (a) Concept and workflow of the Reaction Norm framework. (b) Concept and workflow of the PEI framework.

Prediction accuracy for multiple traits in a maize multi-parental advanced inter-cross population multi-environment trial

We first benchmarked the prediction performance using a maize multi-parental advanced inter-cross population (CUBIC) [29, 46]. This population was derived from 24 elite Chinese maize inbred lines from four divergent heterotic groups, and a total of 24 founders were crossed under a complete diallel cross mating design, omitting reciprocal crosses. A total of 1404 inbred maize lines were obtained and phenotyped at five locations in China’s major maize production zone, covering longitudinal variation from E 114°01’ at Henan (HN) to E 125°18’ at Jilin (JL) and

latitudinal variation from N 43°42’ at JL to N 35°27’ at HN (Fig. 2a) [47]. Furthermore, continuous monitoring of environmental data at the planting locations was conducted (Fig. 2b). Here, we assessed the prediction performance using three traits: days to anthesis (DTA), plant height (PH), and ear weight (EW) with distinct genetic architecture (Fig. S2, Table S1).

We applied the CERIS algorithm to search for an environmental index for each trait. Daily highest temperature (DHT) explained the largest amount of variation for DTA and PH, while daily lowest temperature (DLT) was correlated with EW with the highest correlation (Fig. 2c, Figs S3, S4 Table S2). These indexes were integrated into the two frameworks for predicting these traits across multiple environments. PEI framework exhibited superior prediction accuracy than that from the RN framework, with a 10%, 8.7%,

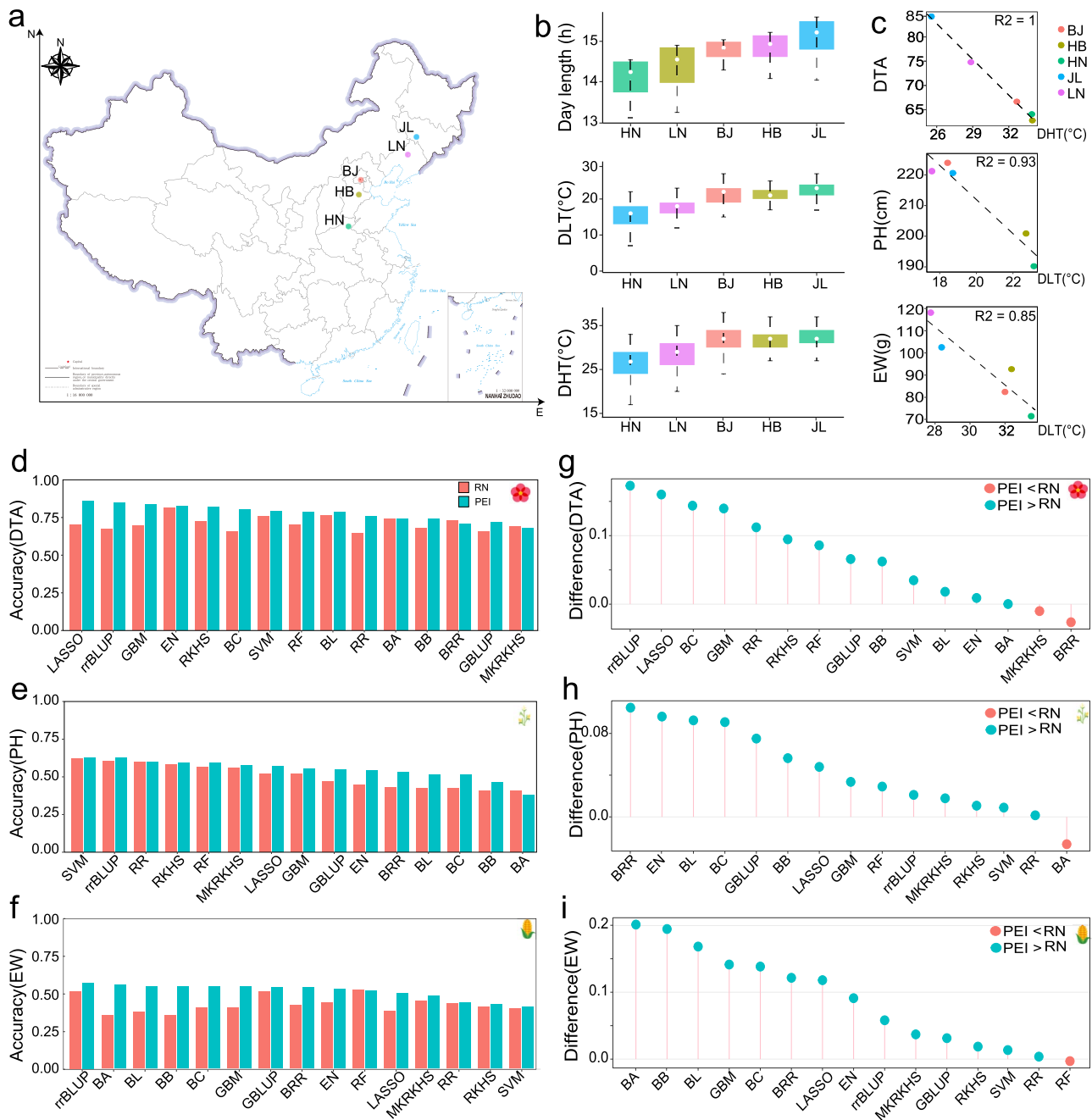


Figure 2 Prediction accuracy for multiple traits in a maize multiple parental advanced inter-cross multi-environment trial. (a) Geographical distribution of the five locations (Map Content Approval Number: GS (2019)1686); (b) Boxplot of the environmental factors at the five locations. (c) Performance of the correlation between the mean phenotypes and environmental index within optimal time windows across diverse locations. (d-f) Performance of the 15 estimators for the two frameworks on Days to Anthesis (DTA), Plant height (PH), and Ear weight (EW); (g-i) Comparison of the prediction accuracy between PEI and RN frameworks across 15 estimators on DTA, PH, and EW.

and 20.6% increase in accuracy for DTA, PH, and EW (Fig. 2d-f, Table S2-S4). It is noteworthy that the optimal estimators varied between different frameworks. Specifically, for DTA, EN yielded the highest prediction accuracy for RN framework, whereas LASSO exhibited superior performance for PEI framework (Fig. 2d). In contrast, for PH, SVM outperformed other estimators for RN and PEI frameworks (Fig. 2e). In the case of EW, rr-BLUP demonstrated the highest accuracy for PEI framework, while RF achieved optimal performance for RN framework (Fig. 2f). Overall, most estimators for the PEI framework demonstrated improved prediction accu-

racy compared to those for the RN framework in this population (Fig. 2g-i).

Prediction accuracy for multiple traits in an inbred wheat population multi-environment trial

We then benchmarked the prediction performance using an inbred wheat population from the International Maize and Wheat

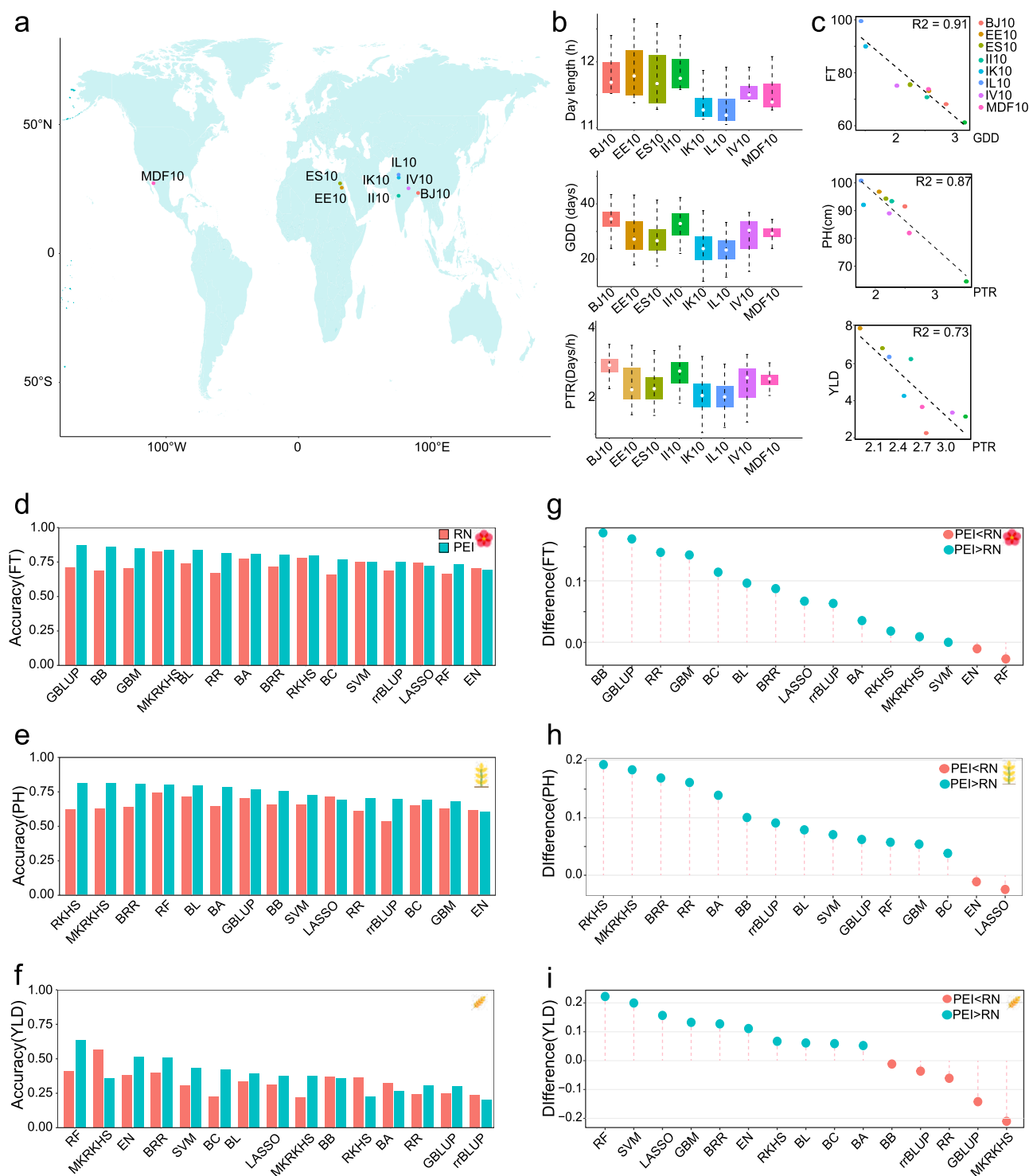


Figure 3 Prediction accuracy for multiple traits in an inbred wheat multi-environment trial. (a) Geographic distribution of experimental locations for the nine locations; (b) Boxplot of the environmental factors at the nine locations. (c) Performance of the correlation between the mean phenotypes and environmental index within optimal time windows across diverse locations. (d-f) Performance of the 15 estimators for the two frameworks on Flowering time (FT), Plant height (PH), and Yield (YLD); (g-i) Comparison of the prediction accuracy between PEI and RN frameworks across 15 estimators on FT, PH, and YLD.

Improvement Center (CIMMYT). This population is part of the Wheat Association Mapping Initiative (WAMI) and was assembled from elite wheat nurseries distributed through the International Wheat Improvement Network (IWIN) [28, 48]. A total of 287 inbred wheat lines were genotyped and phenotyped at nine

locations across major global production zones, including India in Asia, Mexico in North America, and Egypt in North Africa. These locations covered a wide range of geographical regions, from E 90°13' at Joydebpur (BJ10) to W 109°33' at Sonora (MDF10), and from N 15°15' at Dharwad (ID10) to N 30°32' at Ludhiana

Fig. 3a(IL10) (Fig. 3a). Additionally, the environmental data at the planting locations were continuously monitored (Fig. 3b). Here, we evaluated prediction performance using three traits: flowering time (FT), PH, and yield (YLD) with distinct genetic architecture (Fig. S5, Tables S3, S5).

We initially used CERIS to search for an environmental index for each trait. Growing degree days (GDD) explained the largest amount of variation for FT, while the photothermal ratio (PTR) showed the highest correlation with PH and YLD (Fig. 3c, Figs S6, S7). These indexes were integrated into the two frameworks for predicting these traits across multiple environments. Overall, PEI framework exhibited superior prediction accuracy compared to that from the RN framework, with an increase of 10%, 13.9%, and 14.7% for FT, PH, and YLD (Fig. 3d–f, Table S4 Tables S6–S8). Consistent with observations from the CUBIC maize population, optimal estimators varied between two frameworks for different traits. For FT, MKRKHS yielded the highest prediction accuracy for RN framework, whereas G-BLUP outperformed the remaining estimators for PEI framework (Fig. 3d). When it comes to PH, RF outperformed other estimators for RN framework, while the RKHS was the best for PEI framework (Fig. 3e). In the case of YLD, the RF showed the highest accuracy for PEI framework (Fig. 3f). Overall, most estimators for the PEI framework demonstrated higher prediction accuracy than those from the RN framework (Fig. 3g–i).

Prediction accuracy for multiple traits in an F₁ willow population multi-environment trial

We finally benchmarked the prediction performance using an F₁ willow population from a multi-environment trial. This population was derived from a complete diallelic cross among five elite male and five female outbred willow lines [49, 50]. A total of 501 F₁ willow lines were obtained and phenotyped at three locations in Sichuan province, China. These locations ranged longitudinally from E 103°51' at Leshan (LS) to E 104°36' at Yibing (YB), and latitudinally from N 28°47' at YB to N 29°37' at LS (Fig. 4a). Temperature and relative humidity (RH) of the three planting locations were continuously monitored (Fig. 4b). Here, we tested the prediction performance using three traits: defoliation time (DT), PH, and ground diameter (GD), with varying genetic architecture (Fig. S8, Table S5).

Initially, we utilized CERIS to identify an environmental index corresponding to each trait. Daily lowest temperature (DLT) accounted for the greatest variation in DT and PH, while RH exhibited the strongest correlation with GD (Fig. 4c, Figs S9, S10, Table S6). These indexes were integrated into the two frameworks for predicting these traits across multiple environments. While PEI framework consistently outperformed RN across nearly all traits in the two inbred populations, its average accuracy was lower in this outbred population. Compared to the RN framework, PEI framework exhibited an average reduction in accuracy of 20% for PH and 1.6% for GD (Tables S10–S12). However, the PEI framework demonstrated an advantage in prediction accuracy, achieving 20.6% for DT. Furthermore, while most estimators favoured the RN framework over the PEI framework, the PEI framework showed modest prediction accuracy improvement of 9% for DT and 0.9% for GD, when using the best-performing estimators (Fig. 4d–f).

Predicting phenotypes in novel environments

We finally evaluated the prediction performance of known genotypes in a novel environment. To do this, one of the tested environments was treated as unknown, and all phenotype measurements in that environment were set to 'NA'. Prediction accuracy was calculated as the correlation between predicted and observed phenotypes at each site, and these correlations were then averaged across all sites. As the RN framework requires a minimum of three environments for linear regression, the F1 willow population was excluded from this analysis. To address differences in the number of environments between the maize and wheat inbred populations, we down sampled the wheat population to five planting locations.

For the maize CUBIC population, the PEI framework achieved higher predictive accuracy in 53.3% (8 out of 15) of predictions, spanning three traits across five locations, with an average improvement of 16.5% compared to the RN framework. Except for JL, the PEI framework delivered superior accuracy for DTA at all the remaining locations compared to the RN framework, with an average improvement of 13.7% in accuracy (Fig. 5a–e). For PH, the PEI framework consistently outperformed the RN framework at all five locations, with an average improvement of 21.7% in prediction accuracy (Fig. 5f–i). Regarding EW, the PEI framework exhibited higher accuracy for all locations except BJ, with an average improvement of 11.5% (Fig. 5k–o). Overall, the PEI framework demonstrated superior performance compared to the RN framework for predicting phenotypes in novel environments within the maize inbred populations (Fig. 5).

For the inbred wheat population, the PEI framework achieved higher predictive accuracy in 29.1% (9 out of 15) of predictions, encompassing three traits across five locations. However, the RN framework presented an average accuracy improvement of 2.9% over the PEI framework, driven by significantly better performance in the remaining locations (Fig. S11).

Comparison of computing time and memory usage

We finally compared the computing time and memory usage in relation to population size and parameter estimators. Since PEI framework involves the creation of super-matrices, whose dimension scales linearly with the number of environments, it requires more computational resources than RN framework when the number of environments increases. Here, we evaluated the computational cost under different population sizes by down and up sampling the CUBIC maize population for both frameworks and all the parametric estimators (Fig. 6a–d). For computational time, GBM has the shortest runtime, while the BL required the longest computing time. Meanwhile, GBM utilized the least amount of memory for both frameworks, while RF consumed the most. In addition, computing time and memory usage do not scale in a uniform manner across estimators (Fig. 6a–d). All the estimators exhibit an increase in computation time and memory usage as the data dimension increases (Fig. 6e and f). Semi-parametric estimators (e.g. RKHS and MKRKHS) showed the greatest increase in computing time and memory usage, while non-parametric estimators (e.g. GBM and SVM) remained relatively stable as the data

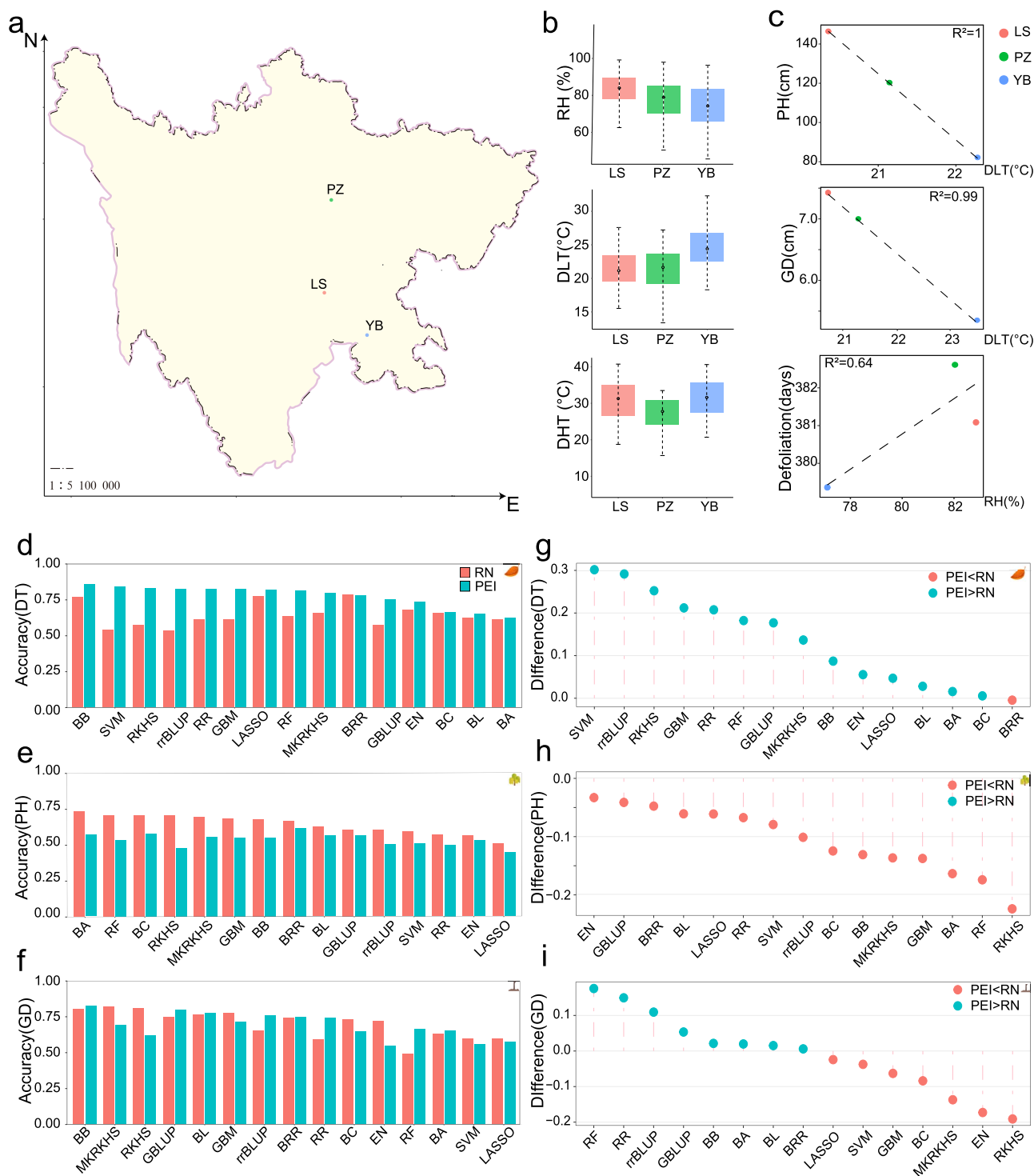


Figure 4 Prediction accuracy for multiple traits in an F1 willow multi-environment trial. (a) Geographic distribution of experimental locations for willow lines (Map Content Approval Number: GS (2017)1267); (b) Boxplot of the environmental datasets at the three locations. (c) Performance of the correlation between the mean phenotypes and environmental index within optimal time windows across diverse locations. (d-f) Prediction accuracy of the 15 estimators for the two frameworks on DT: Defoliation time; PH: Plant height; GD: Ground diameter; (g-i) Comparison of the prediction accuracy between PEI and RN frameworks across 15 estimators on DT, PH, and GD.

dimensionality increases (Fig. 6e and f). Running PEI on a standard workstation (e.g. 16-core CPU, 64 GB RAM) showed that when dimension increases from 7 to 8, the computational cost increased substantially while prediction accuracy remained low (Fig. S12). Here, ‘dimension’ refers to the order of magnitude (base 10) of the

product of population size and number of markers. For example, 100 lines \times 10 000 markers = 1 000 000 = 10^6 , giving a dimension of 6. When dimension increases from 7 to 8, the gain in computational cost is high, but the accuracy is low. Therefore, users are advised to consider alternative estimators under this scenario.

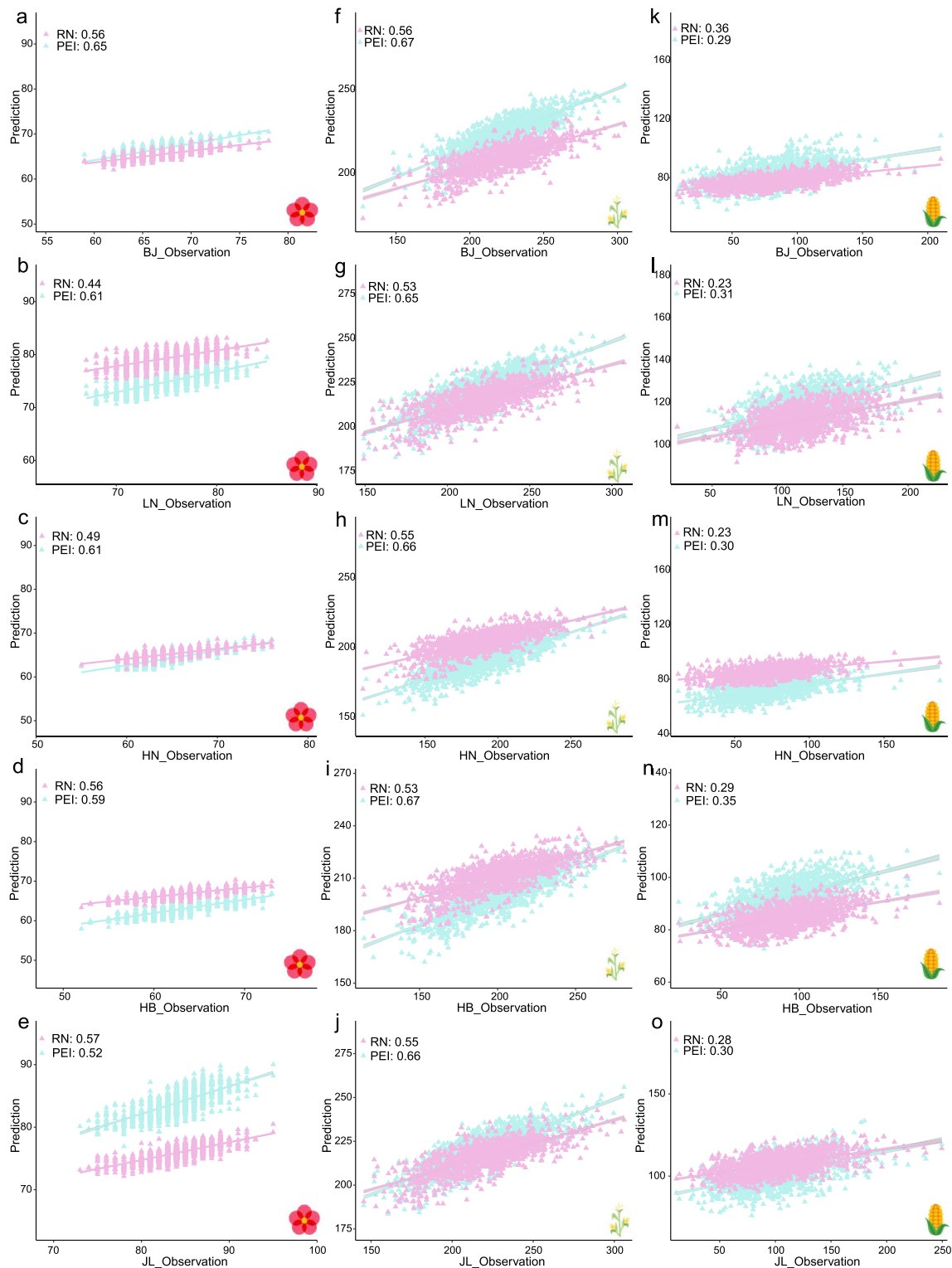


Figure 5 Predicting multiple traits in novel environments for a maize multi-environment trial. (a-e) The forecasting of DTA in the maize CUBIC METs for 5 locations (BJ, LN, HN, HB, JL); (b-j) The forecasting of PH in the maize CUBIC METs for 5 locations; (k-o) The forecasting of EW in the maize CUBIC METs for 5 locations.

Discussion

In this study, we presented a novel framework to model complex traits variation in METs by directly accounting for the combined effects from genotype, environment, and their

interactions. By incorporating a super matrix that integrates environmental index effects and their interactions with the polygenic background, this approach significantly improved our capacity to model complex traits in METs (Fig. 1b). To mitigate the computational challenges associated with handling

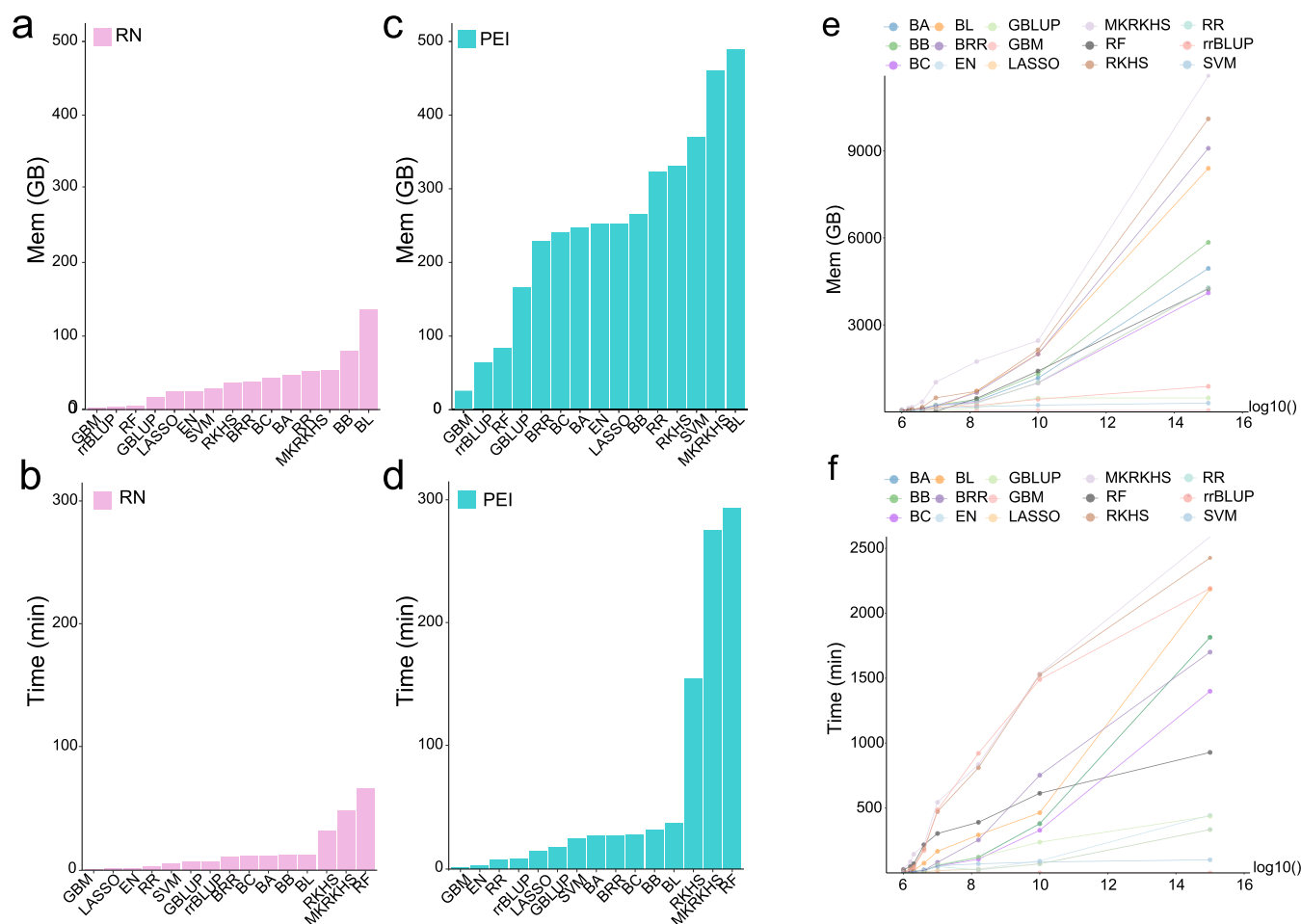


Figure 6 Comparison of computing time and memory usage. (a-b) Memory consumption and computing time of the 15 estimators for RN framework; (d-c) Memory consumption and computation time of the 15 estimators for PEI framework; (e-f) Scaling of memory consumption and computation time in relation to data dimension. Dimension (X-axis) was calculated as sample size multiplied by the number of genotype markers. They were transformed to -log10 scale to increase readability.

a massive number of markers, we adopted two solutions. One involved marker pruning based on linkage disequilibrium [51]. The other leveraged machine learning and deep learning approaches to maintain prediction accuracy without significantly increasing computational burden. These approaches allowed our framework to match the performance of existing estimators. Furthermore, it did so without incurring an excessive computational load.

Through benchmarking analysis with multiple traits across diverse populations, we observed that the PEI framework significantly outperforms the RN framework in two inbred populations (Figs 2-4). However, in the F₁ willow population, the PEI framework showed superior predictive ability only for DT. This is likely because complex interactions associated with various traits in a hybrid population (Fig. 4). Previous studies in hybrid populations have revealed that epistatic variance is the major component of phenotypic variance [52, 53]. We therefore tested whether RN framework is better at capturing such non-additive effects using a yeast growth dataset. This dataset contains 4000 genotype and phenotype measurements of yeast growth on 20 different medium, and four early studies have mapped out the additive QTLs, and epistatic network in 20 environments [52] (Supplementary Notes). We found that the

slope parameter from the RN framework was correlation with the degree of epistatic interaction, whereas the intercept parameter did not (Fig. S13). This suggests that the RN slope parameter could capture epistatic interactions indirectly. In contrast, the PEI framework may be less suited to model the complex genetic architecture in hybrid populations, where epistasis contributes substantially to phenotypic variation. Our findings highlight the profound influence of genetic architecture and population structures on prediction performance. We, thus, recommend performing comprehensive comparisons to select the optimal strategy in future applications.

It should be noted that predicting phenotypes in novel environments remains challenging due to the complex gene-environment interactions. This challenge is particularly pronounced for METs that span a wide range of environments. In the inbred wheat METs analysed here, the PEI framework performed poorly for IH10, likely because IH10 differs significantly from the other planting locations (Fig. 3 and Figs S8, S14). Distance-based analyses revealed that IH10 exhibits both the largest minimum distance and the greatest overall dissimilarity relative to the remaining environments, indicating a weak environmental continuity (Fig. S14). This discontinuity reduces prediction performance in the PEI framework, which assumes the pattern of genotype by environment interaction is

similar across tested sites. Additional analysis showed that prediction accuracy is consistently reliable when an environment's Youden's index to the training set is ≤ 3.75 . Above this threshold, prediction accuracy drops substantially (Supplementary Notes, Fig. S14). Thus, the environmental outlier (IH10) reduces prediction accuracy in the PEI framework due to weak environmental continuity (Fig. S14). Increasing the environmental continuity of planting locations might capture the complex interaction pattern between a larger range of ecological factors and genotype, but it would certainly increase the costs of METs [54].

Finally, both RN and PEI frameworks include genetic interaction with the single most relevant environmental index, which introduces two potential caveats. Model performance might be sensitive to the selected environment index and including multiple indexes might improve the model performance. Additional analysis shown that compared to the RN framework, PEI framework exhibited substantially low sensitivity to environmental variation (Fig. S15). These indicate that the PEI is less sensitive than the RN to selection of environmental index, which yielded more robust performance. Furthermore, we demonstrated that while model complexity and computational time increased significantly with including more environment index, prediction accuracy improved in only a few trait and population combinations (Fig. S16). This is because most of the phenotypic variance is captured by the first environmental index, making the gain from adding further index marginal (Fig. S17). In addition, we attempted to benchmark our implementation with existing frameworks with publicly available code. In contrast to our framework, where CERIS always determines the optimal factors, GEFormer used all available environmental data as input, allowing the framework to integrate it independently, which to some extent avoids bias. Despite this, we found no or only marginal gain in prediction accuracy for various trait and site combinations, suggesting our PEI framework is relatively robust (Table S13 and Figs S18, S19). However, our attempt to benchmark against a recently published framework. AutoGS [55], failed due to unstable release of the software going through constant updates. A future comparison will be essential for a fully comprehensive evaluation once a stable, well-documented version of the software is released.

In summary, we introduce a unified framework, PEI, that directly models genotype-by-environment interactions through integrating genotypes and ECs. We implemented 15 estimators, including parametric, non-parametric, and machine learning approaches, to improve prediction accuracy. Alongside the classical RN framework and recently developed deep learning approaches, we conducted benchmarking analysis using three distinct populations with various population structures and three traits with variable genetic architectures. Additionally, we released an R package, MMGS available on GitHub (<https://github.com/Ryougi-yukiro/MMGS>). Our framework and software package could facilitate multi-environment genomic prediction, improve breeding efficiency, and offer deeper insights into modelling the relationship between genotype and environment.

Materials and methods

Datasets used for genomic prediction

Three datasets, reflecting diverse population sizes and structures, were utilized in this investigation. The first dataset is CUBIC maize

population, including 1404 inbred maize lines, and their performance was assessed at five locations within China's primary maize production zone. This zone exhibits longitudinal variation ranging from E 114°01' at HN to E 125°18' at JL and latitudinal variation from N 43°42' at JL to N 35°27' at HN [46]. In essence, this population was derived from 24 elite varieties, representing four divergent heterotic groups, through cycles of random mating, selection, and inbreeding. Three environmental variables, specifically DHT and DLT, along with daily length (DL), were downloaded from Zan *et al.* Genotyping data of the CUBIC maize population were a subset of the maize variants file uploaded, then we filtered these data through Plink 1.9 [56]. A total of 28 973 SNPs among 1007 individuals from 5 locations were used for the next analysis [47].

The second dataset, the inbred wheat population downloaded from the CIMMYT [2], consists of 287 inbred lines planted in India, Pakistan, Nepal, Bangladesh, Iran, Egypt, Sudan, and Mexico. In each location, plants were grown in plots with an α -lattice design with two replications. Four environmental variables, specifically GDD, PTR, and photothermal time (PTT), along with DL, were downloaded in Li *et al.* [22]. Each genotype was planted in a single-row plot of 8 to 15 plants. The wheat population was genotyped with an Illumina iSelect 90K SNP assay for 26 814 SNPs [48]. The HapMap file was formatted into plink format through Tassel [57].

The third dataset, an F1 willow population, consisted of 501 F₁ lines planted in YB, Pengzhou (PZ), and LS. Each genotype was planted in a single-row plot of a single plant. Three environmental variables, specifically DHT and DLT, along with RH were collected from three plots. Whole genome sequencing data of the willow population were processed by GATK v3.8, and over 12 million willow SNPs with missing rate higher than 20% and minor allele frequency less than 1% were obtained. Then all variants were imputed by Beagle [58] following the default parameters.

DNA sequencing and environmental factor acquisition

Total DNA was collected and extracted from fresh leaves of *Salix suchowensis* using the CTAB method. Libraries were constructed and sequencing were performed by Novogene using Illumina NovaSeq platform. Raw reads were quality controlled through fastp [59]. Then, we used bwa-mem2 to map raw reads to the reference genome. The GATK pipeline was utilized for SNP calling [60]. SNPs were subsequently filtered using VCFtools to retain only bi-allelic variants with call rates above 90%, and minor allele frequency greater than 0.05 [61].

To obtain the environmental factors, continuous monitoring was carried out using micro weather stations, built by the Beijing Aozuo Ecological Instrument Co., set up in the field from May to November each year. The system consists of various types of meteorological sensors, data acquisition system, power supply system and so on. It measures wind speed and direction (10m, 5m, 2m), temperature and RH (10m, 5m, 2m), rainfall, atmospheric pressure, gross radiation, net radiation, soil moisture, soil temperature and electrical conductivity at regular intervals and automatically according to the measurement intervals set by the user. The monitoring data excludes the days missed by the machine and the final environmental data is used for the time for which continuous records exist at all three locations.

CERIS for searching environmental index

The CERIS algorithm used for identifying the environmental indexes had been described in the previous study [22]. For the tested populations, we began by calculating the mean values for all individuals across various phenotypes, which were then used as the horizontal coordinates. In addition, we conducted sliding window calculations, utilizing both environmental and phenotypic data over a continuous observation period with equal-sized bins. This procedure allowed us to identify the most appropriate start and end times for corresponding traits.

To validate the relevance of this approach, we performed a regression analysis. This involved using environmental data from various sites within the identified window as explanatory variable, while the mean trait values from these sites were used as the response variable. The environmental factor that yielded the highest R^2 value was considered as the most relevant factor and was selected for further analysis. This approach ensured that the identified window and environmental factor were robustly linked to the phenotypic traits under consideration.

Reaction norm framework

Construction of the RN framework relies heavily on previous research [18, 22]. It quantifies individual response to changes in environment factor as a unique intercept and slope using linear regression. For each individual i , a linear regression was performed with the environmental index detected in previous section as the explanatory variable x and the phenotypes in environment j as response variable y_{ij} .

$$y_{ij} = k_i x + b_i + e_{ij} \quad (1)$$

y_{ij} represents the phenotypes of i individual in j environment, x is the environmental-index value of j environment centred by the average environmental-index values across the environments, and e_{ij} is the residual. k_i and b_i is the slope and intercept for individual i , representing two unique traits of environmental response.

Genomic prediction in the RN framework using rr-BLUP was performed as:

$$y = \mu + Zg + \varepsilon \quad (2)$$

where y is the phenotypic values (i.e. either intercept or slope derived for equation (1)), μ is the overall mean. g is a vector of random genetic effects, and ε is a vector of residuals variance. Z are incidence matrices. The variance of the random effects g is $\text{var}(g) = G\sigma_g^2$, where G is the genomic relationship matrix and σ_g^2 is the additive genetic variance. Predicted phenotype for a specific individual was obtained by plugging in estimated k_i and b_i from equation (2) back to equation (1).

Polygenic-environment-interaction framework

We modelled the phenotype measured in multiple environments using mixed linear model.

Let y_{ij} be the phenotype of i th ($i = 1, 2, 3, \dots, n$) individual measure at j th ($j = 1, 2, 3, \dots, m$) environment. In a specific environment j , y_{ij} (length n) is affected by a linear combination of random marker

effects u_j and residual error e_j . The length of u_j is equal to the number of markers, p , and the random marker effects are assumed identically and independently distributed (iid) and normal, so that $u_j \sim N(0, I\sigma_u^2)$. e_j is the residual error $e \sim N(0, I\sigma_e^2)$. Thus, a linear mixed model for this environment is:

$$y_j = \mu_j + Mu_j + e_j \quad (3)$$

whereby μ is a population mean, and matrix M is the additive genotype matrix coded as 0, 1, and 2, with n by p dimension. n is the number of individuals and p is the number of markers

$$M = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 2 & 1 & \dots & 2 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 2 & \dots & 1 \end{pmatrix} \quad (4)$$

An equivalent G-BLUP model can be obtained by define an individual random effect α_j (length n). such that $\alpha_j = Z_j u_j$, and we thereby have:

$$y_j = \mu_j + \alpha_j + e_j \quad (5)$$

where $V(\alpha) = G\sigma_\alpha^2$. Here G is the genomic relationship matrix, and we have $G = Z_j Z_j'$.

Extending this univariant mixed model to multiple environments, we could construct a multi environment phenotype vector y by stacking up the phenotype vector, y_1 to y_m , measured from each environment.

$$y_j = \begin{pmatrix} y_{1j} \\ y_{2j} \\ \vdots \\ y_{nj} \end{pmatrix}, \quad y = \begin{pmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{nm} \end{pmatrix} \quad (6)$$

Assuming there is no genotype by environment interaction, we have

$$y = \mu + E + Bu + e \quad (7)$$

whereby E is an environment specific environmental effect vector (length $n \times m$), u and e is the stacked marker effects and residual errors, and B is a super genotype matrix with $n \times m$ rows and p columns.

$$E = \begin{pmatrix} E_1 \\ E_2 \\ \vdots \\ E_m \end{pmatrix} \quad (8)$$

E_j is a vector of length n representing the most relevant environmental effects, which is obtained using the CERIS algorithms as described above.

$$B = I_m \otimes M = \begin{pmatrix} M & 0 & \dots & 0 \\ 0 & M & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & M_m \end{pmatrix} \quad (9)$$

I_m is an identity matrix with m dimension, which corresponds to the number of environments. B is a diagonal matrix with genotype matrix, excluding E .

$$u = \begin{pmatrix} u_{11} \\ u_{21} \\ \vdots \\ u_{pm} \end{pmatrix} \quad (10)$$

$$e = \begin{pmatrix} e_{11} \\ e_{21} \\ \vdots \\ e_{nm} \end{pmatrix} \quad (11)$$

An equivalent G-BLUP model can be obtained by define an individual random effect α (length $n \times m$). such that

$$\alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{pmatrix} \quad (12)$$

and we thereby have:

$$y = \mu + E + Z\alpha + e \quad (13)$$

where $V(\alpha) = K\sigma_u^2$. Here, $K = I_m \otimes \mathbf{G}$ is the relationship matrix. And this model can be solved as a univariate G-BLUP model.

To model the interaction between the environment and polygenic background, we constructed a genotype by environment interaction matrix at environment j with the following equation:

$$M_j^{\text{inter}} = E_j M = \begin{pmatrix} 1E_j & 0E_j & \cdots & 0E_j \\ 2E_j & 1E_j & \cdots & 2E_j \\ \vdots & \vdots & \ddots & \vdots \\ 0E_j & 2E_j & \cdots & 1E_j \end{pmatrix} \quad (14)$$

We then combined the matrix M and M_j^{inter} as a new genotype matrix M_j^{all}

$$M_j^{\text{all}} = [M \ M_j^{\text{inter}}] \quad (15)$$

This is equivalent to double the number of makers by including a pseudo marker vector representing the interaction from each of these markers and the environment index.

Similarly, a multiple environment super genome matrix could be constructed.

$$B_{\text{all}} = \begin{pmatrix} M_1^{\text{all}} & 0 & \cdots & 0 \\ 0 & M_2^{\text{all}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & M_m^{\text{all}} \end{pmatrix} \quad (16)$$

where B_{all} is a diagonal matrix with new genotype matrix M_j^{all} combined with E_j .

Therefore, by plugging this back to model (7) and (13), we could make predictions for multi-environment phenotype in the mixed linear model framework.

Benchmark against GEFormer framework

We followed the original design of GEFormer and applied it to our datasets. Hyperparameter selection was performed using the Optuna Bayesian optimization library, with a search conducted on the training set for various traits of three populations to obtain optimal configuration. All GEFormer analyses were implemented in Python 3.9.19 using PyTorch. A complete description of training hyperparameters across various datasets in the supplementary tables (Table S14).

Cross-validation schemes

PEI performance predicted phenotypes encompass two scenarios: (i) forecasting the performance of untested genotypes in tested environments, and (ii) predicting the performance of tested genotypes in untested environments. Unlike traditional joint regression analyses, such as the RN framework, the PEI framework relies on genotypic data and exploits potential relationships between tested genotypes and phenotypes to predict the phenotypes of untested genotypes. Additionally, by analysing correlations between environments and environmental indices, the method forecasts the performance of individuals without phenotypic data in future environments, drawing upon phenotypic data from existing environments.

In the first scenario, a leave-one-half-of-genotypes-out cross-validation was performed. (i) The n genotypes were randomly divided into two equal groups: tested genotypes and untested genotypes. (ii) The environmental index was identified using CERIS, based on the environmental means derived from the tested genotypes. (iii) Model (13) was then used to make predictions for the untested genotypes. Genomic prediction was then executed using various estimators to forecast the phenotypes of each untested genotype.

In the second scenario, leave-one-environment-out cross-validation was conducted. (i) Each environment was sequentially treated as the untested environment, with the remaining environments serving as the tested environments. (ii) The environmental index was identified using CERIS, based on the environmental means of the tested genotypes in the tested environments. (iii) With the obtained environmental index, a genotype by environment interaction matrix was built using model (16). (iv) Model (13) was then used to make predictions for the untested genotypes.

Prediction accuracy was assessed as the correlation between observed and predicted values. The predicted values were derived from a framework built using the training set of tested genotypes in tested environments. Accuracy was evaluated both across all environments and at the individual environment level. During genotype sampling, the average prediction accuracy across five replications with 10-fold was calculated. A representative run is provided in the figure for illustration.

Description of modelling and parameter estimators

We employed the *sf* [62] package to generate a map of China and the *rworldmap* [63] package to create a world map. Consider that forecasting estimators are divided into parametric and

non-parametric estimators and that the principles of different parametric estimators vary. Most of them are based on additive models, and their accuracies may be different because they vary in their assumptions and algorithms with respect to the variances of complex traits. Incorporating non-additive effects or multiple variates, the general estimators can be extended (Fig. S20) [15, 64]. Various estimators were implemented by leveraging parameters from R packages [42, 43, 65, 65–67]. Finally, we illustrated the average prediction accuracy of 15 distinct estimators, encompassing parametric, semi-parametric, and non-parametric approaches, for both frameworks using ggplot2 [68], with adjustments made in Adobe Illustrator.

Heritability estimation

The kinship heritability of each trait was calculated as follows:

$$H = \frac{V_g}{V_g + \frac{V_{ge}}{L} + \frac{V_e}{RL}} \quad (17)$$

where V_g the epigenetic variance, V_{ge} represents the gene-environment interaction variance, L indicates the number of environments, V_e signifies the environmental variance, and R was the number of replicates.

Acknowledgements

This research was funded by National Key Research and Development Program of China (2021YFD2200202), the State Key Research & Development Project-Youth Scientist program (2023YFD1202400), National Science Foundation of China (32200503), Taishan Young Scholar Program and Distinguished Overseas Young Talents Program from Shandong province (2024HWYQ-079).

Author contributions

J.L. and Y.Z. conceived and supervised this project and its components. Y.H., W.L., Z.Z., and M.Z. collected raw datasets from open-source websites. Y.H., W.L., and M.Z. conducted the experiments. X.D. and H.W. built the hybrid willow populations. W.M., M.Z., and Y.Z. performed the development of this R package and M.Z. presented the data analysis. The plant locations: PZ (pengzhou), LS (leshan), YB (yibing), were chosen by J.L. M.Z., J.L., and Y.Z. wrote the manuscript.

Data availability

The test datasets, including genotypic, phenotypic, and environmental data, can be accessed at this GitHub repository (<https://github.com/Ryougi-yukiro/MMGS-bench>). Additionally, the benchmark codes have been uploaded to this repository. For quick installation of the developed R package, kindly refer to the GitHub repository: <https://github.com/Ryougi-yukiro/MMGS>. This repository not only facilitates the installation of the R package but also provides example codes tailored for quick start. Detailed instructions can be found in the R documents and accompanying tutorial (https://multiplemethods.gitbook.io/MMGS_tutorial_v1).

Conflicts of interest statement

No conflict of interest was declared.

Supplementary material

Supplementary material is available at *Horticulture Research* online.

References

1. Scheiner SM. Genetics and evolution of phenotypic plasticity. *Annu Rev Ecol Syst.* 1993;24:35–68
2. Rajaram S, Borlaug NE, Van Ginkel M. CIMMYT international wheat breeding. In: *Bread Wheat Improvement and Production*. FAO: Plant Production and Protection Series, 2002,30:103–17
3. Kelly M. Adaptation to climate change through genetic accommodation and assimilation of plastic phenotypes. *Philos Trans R Soc B.* 2019;374:20180176
4. Nationen V. *World Population Prospects 2022: Summary of Results*. New York, UN; 2022.
5. Hickey LT, Hafeez AN, Robinson H. *et al.* Breeding crops to feed 10 billion. *Nat Biotechnol.* 2019;37:744–54
6. Yu J, Pressoir G, Briggs WH. *et al.* A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet.* 2006;38:203–8
7. Peng B, Guan K, Tang J. *et al.* Towards a multiscale crop modelling framework for climate change adaptation assessment. *Nat Plants.* 2020;6:338–48
8. Crossa J. Statistical analyses of multilocation trials. *Adv Agron.* 1990;44:55–85
9. Huehn M. Nonparametric measures of phenotypic stability. Part 1: theory. *Euphytica.* 1990;47:189–94
10. Yan W, Tinker NA. Biplot analysis of multi-environment trial data: principles and applications. *Can J Plant Sci.* 2006;86: 623–45
11. Burdon RD. Genetic correlation as a concept for studying genotype-environment interaction in forest tree breeding. *Silvae Genet.* 1977;26:168–175.
12. Yan W, Kang MS, Ma B. *et al.* GGE Biplot vs. AMMI analysis of genotype-by-environment data. *Crop Sci.* 2007;47:643–53
13. Malosetti M, Ribaut J-M, van Eeuwijk FA. The statistical analysis of multi-environment data: modeling genotype-by-environment interaction and its genetic basis. *Front Physiol.* 2013;4:37433
14. Cullis BR, Jefferson P, Thompson R. *et al.* Factor analytic and reduced animal models for the investigation of additive genotype-by-environment interaction in outcrossing plant species with application to a *Pinus radiata* breeding programme. *Theor Appl Genet.* 2014;127:2193–210
15. Crossa J, Pérez-Rodríguez P, Cuevas J. *et al.* Genomic selection in plant breeding: methods, models, and perspectives. *Trends Plant Sci.* 2017;22:961–75
16. Kolmodin R, Bijma P. Response to mass selection when the genotype by environment interaction is modelled as a linear reaction norm. *Genet Sel Evol.* 2004;36:435
17. Pierce BA. *Genetics: A Conceptual Approach*. Macmillan Learning, 2020.

18. Jarquín D, Crossa J, Lacaze X. *et al.* A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theor Appl Genet.* 2014;127:595–607
19. Jarquín D, Howard R, Crossa J. *et al.* Genomic prediction enhanced sparse testing for multi-environment trials. *G3 (Bethesda).* 2020;10:2725–39
20. Yu X, Li X, Guo T. *et al.* Genomic prediction contributing to a promising global strategy to turbocharge gene banks. *Nat Plants.* 2016;2:1–7
21. Jarquín D, Lemes da Silva C, Gaynor RC. *et al.* Increasing genomic-enabled prediction accuracy by modeling genotype × environment interactions in Kansas wheat. *Plant Genome.* 2017;10. <https://doi.org/10.3835/plantgenome2016.12.0130>
22. Li X, Guo T, Wang J. *et al.* An integrated framework reinstating the environmental dimension for GWAS and genomic selection in crops. *Mol Plant.* 2021;14:874–87
23. Smith AB, Ganesalingam A, Kuchel H. *et al.* Factor analytic mixed models for the provision of grower information from national crop variety testing programs. *Theor Appl Genet.* 2015;128:55–72
24. Gamal El-Dien O, Ratcliffe B, Klápště J. *et al.* Prediction accuracies for growth and wood attributes of interior spruce in space using genotyping-by-sequencing. *BMC Genomics.* 2015;16:370
25. Crowell S, Korniliev P, Falcão A. *et al.* Genome-wide association and high-resolution phenotyping link *Oryza sativa* panicle traits to numerous trait-specific QTL clusters. *Nat Commun.* 2016;7:10527
26. Li M, Zhang Y-W, Zhang Z-C. *et al.* A compressed variance component mixed model for detecting QTNs and QTN-by-environment and QTN-by-QTN interactions in genome-wide association studies. *Mol Plant.* 2022;15:630–50
27. Zhang Z, Ersoz E, Lai C-Q. *et al.* Mixed linear model approach adapted for genome-wide association studies. *Nat Genet.* 2010;42:355–60
28. Sukumaran S, Lopes M, Dreisigacker S. *et al.* Genetic analysis of multi-environmental spring wheat trials identifies genomic regions for locus-specific trade-offs for grain weight and grain number. *Theor Appl Genet.* 2018;131: 985–98
29. Liu H-J, Wang X, Xiao Y. *et al.* CUBIC: an atlas of genetic architecture promises directed maize improvement. *Genome Biol.* 2020;21:20
30. Zan Y, Carlborg Ö. Dynamic genetic architecture of yeast response to environmental perturbation shed light on origin of cryptic genetic variation. *PLoS Genet.* 2020;16: e1008801
31. Ji Y, Han Y, Dai Y. *et al.* Polygenic basis of strong and rapid flowering time response to environment perturbations in wild *Arabidopsis thaliana* population. *bioRxiv.* 2023.
32. Alemu A, Åstrand J, Montesinos-López OA. *et al.* Genomic selection in plant breeding: key factors shaping two decades of progress. *Mol Plant.* 2024;17:552–78
33. VanRaden PM. Efficient methods to compute genomic predictions. *J Dairy Sci.* 2008;91:4414–23
34. Meuwissen TH, Hayes BJ, Goddard M. Prediction of total genetic value using genome-wide dense marker maps. *Genetics.* 2001;157:1819–29
35. George EI, McCulloch RE. Variable selection via Gibbs sampling. *J Am Stat Assoc.* 1993;88:881–9
36. Erbe C, MacGillivray A, Williams R. Mapping cumulative noise from shipping to inform marine spatial planning. *J Acoust Soc Am.* 2012;132:EL423–8
37. Park T, Casella G. The Bayesian Lasso. *J Am Stat Assoc.* 2008; 103:681–6
38. Usai MG, Goddard ME, Hayes BJ. LASSO with cross-validation for genomic selection. *Genet Res.* 2009;91:427–36
39. Whittaker RJ. Scale, succession and complexity in island biogeography: are we asking the right questions? *Glob Ecol Biogeogr.* 2000;9:75–85
40. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc Ser B Stat Methodol.* 2005;67: 301–20
41. Gianola D. Priors in whole-genome regression: the Bayesian alphabet returns. *Genetics.* 2013;194:573–96
42. Maenhout S, De Baets B, Haesaert G. *et al.* Support vector machine regression for the prediction of maize hybrid performance. *Theor Appl Genet.* 2007;115:1003–13
43. Chen X, Ishwaran H. Random forests for genomic data analysis. *Genomics.* 2012;99:323–9
44. Ke G, Meng Q, Finley T. *et al.* Lightgbm: a highly efficient gradient boosting decision tree. *Adv Neural Inf Proces Syst.* 2017;30:
45. Yan J, Xu Y, Cheng Q. *et al.* LightGBM: accelerated genomically designed crop breeding through ensemble learning. *Genome Biol.* 2021;22:271
46. Luo J, Wei C, Liu H. *et al.* MaizeCUBIC: a comprehensive variation database for a maize synthetic population. *Database.* 2020;2020:baaa044
47. Jin M, Liu H, Liu X. *et al.* Complex genetic architecture underlying the plasticity of maize agronomic traits. *Plant. Communications.* 2023;4:100473
48. Sukumaran S, Dreisigacker S, Lopes M. *et al.* Genome-wide association study for grain yield and related traits in an elite spring wheat population grown in temperate irrigated environments. *Theor Appl Genet.* 2015;128:353–63
49. Liu J, Yin T, Ye N. *et al.* Transcriptome analysis of the differentially expressed genes in the male and female shrub willows (*Salix suchowensis*). *PLoS One.* 2013;8:e60181
50. Yan Z, Chen L, Guo Y. *et al.* Pan-transcriptome analysis of willow species from diverse geographic distributions. *Forests.* 2023;14:1182
51. Won S, Park J-E, Son J-H. *et al.* Genomic prediction accuracy using haplotypes defined by size and hierarchical clustering based on linkage disequilibrium. *Front Genet.* 2020; 11:134
52. Forsberg SKG, Bloom JS, Sadhu MJ. *et al.* Accounting for genetic interactions improves modeling of individual quantitative trait phenotypes in yeast. *Nat Genet.* 2017;49: 497–503
53. Liu N, Du Y, Warburton ML. *et al.* Phenotypic plasticity contributes to maize adaptation and heterosis. *Mol Biol Evol.* 2021;38:1262–75
54. Yao Z, Yao M, Wang C. *et al.* GEFormer: A genomic prediction method of genotype-environment interaction in maize by integrating gating mechanism MLP and linear attention mechanism. *Mol Plant.* 2025;13:1135.
55. Wu H, Han R, Zhao L. *et al.* AutoGP: an intelligent breeding platform for enhancing maize genomic selection. *Plant Commun.* 2025;6:101240

56. Purcell S, Neale B, Todd-Brown K. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81:559–75
57. Bradbury PJ, Zhang Z, Kroon DE. *et al.* TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics.* 2007;23:2633–5
58. Pook T, Mayer M, Geibel J. *et al.* Improving imputation quality in BEAGLE for crop and livestock data. *G3 (Bethesda).* 2020;10:177–88
59. Chen S, Zhou Y, Chen Y. *et al.* fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics.* 2018;34:i884–90
60. De Summa S, Malerba G, Pinto R. *et al.* GATK hard filtering: tunable parameters to improve variant calling for next generation sequencing targeted gene panel data. *BMC Bioinformatics.* 2017;18:119
61. Danecek P, Auton A, Abecasis G. *et al.* The variant call format and VCFtools. *Bioinformatics.* 2011;27:2156–8
62. Pebesma EJ. Simple features for R: standardized support for spatial vector data. *R J.* 2018;10:439
63. South A. rworldmap: a new R package for mapping global data. *R J.* 2011;3:3
64. Wang K, Abid MA, Rasheed A. *et al.* DNNGP, a deep neural network-based method for genomic prediction using multi-omics data in plants. *Mol Plant.* 2023;16:279–93
65. Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch F. (2025). e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. R package version 1.7-17, <https://CRAN.R-project.org/package=e1071>.
66. Endelman JB. Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome.* 2011;4:250–5
67. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw.* 2010;33:1–22
68. Wickham H. ggplot2. *WIREs Comp Stats.* 2011;3:180–5